BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/82, 15/29, C12Q 1/68
A01H 5/00, 5/02

A1

(11) International Publication Number: WO 93/14211

(43) International Publication Date: 22 July 1993 (22.07.93)

(21) International Application Number: PCT/GB93/00019

(22) International Filing Date: 8 January 1993 (08.01.93)

(30) Priority data:
818,570 9 January 1992 (09.01.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 818,570 (CIP)
Filed on 9 January 1992 (09.01.92)

(71)(72) Applicants and Inventors: COEN, Enrico [GB/GB]; CARPENTER, Rosemary [GB/GB]; GOODRICH, Justin [GB/GB]; John Innes Institute, Colney Lane, Norwich, Norfolk NR4 7UN (GB).

FREEDMAN, Roger [GB/GB]; Diatech Limited, 19 Millers Yard, 10/11 Mill Lane, Cambridge CB2 1RQ (GB).

(74) Agents: MARCHANT, James, Ian et al.; Elkington and Fife, Prospect House, 8 Pembroke Road, Sevenoaks, Kent TN13 1XR (GB).

(81) Designated States: AU, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: REGULATION OF PLANT GENES

(57) Abstract

A method of regulating the expression of one or more anthocyanin pigment genes in a plant which comprises the steps of transforming plant tissue with an expression vector comprising a DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 and which is capable of regulating expression of one or more plant genes involved in pigment biosynthesis, the said DNA segment being under the control of a promoter upstream of and operably linked thereto and regenerating from the transformed tissue plants showing altered anthocyanin pigmentation.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslowakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

REGULATION OF PLANT GENES

This invention relates to the regulation of plant genes, more particularly the regulation of genes which control the pigmentation of plants.

Technological Background.

Plant species display remarkable diversity in the pattern and intensity of their pigmentation, in particular red or purple anthocyanin pigmentation. Mutations that block anthocyanin production are viable and have readily observable phenotypes; consequently two classes of genes affecting anthocyanin biosynthesis have been characterised in a range of species including maize, *Antirrhinum majus*, pea and *Petunia hybrida* (see Coe and Nuffer in *Corn and corn improvement*, ed. Sprague, 19-53, American Society of Agronomy, Madison, Wisconsin (1977), Dooner et al, *Ann. Rev. Genet.*, 25, 173-199 (1992), Martin et al, *Soc. Exp. Biol. Symp.*, 32, 19-52 (1987), Harker et al, *The Plant Cell*, 2, 185-194 (1990) and Gerats and Martin, Recent Advances in Phytochemistry, H. Stafford, Ed., (in press).

The genes of one class encode enzymes required for pigment biosynthesis and many of these genes appear to be common to different species (see Martin et al, *The Plant Journal*, 1, 37-49 (1991), Sommer and Saedler, *Mol. Gen. Genet.*, 202, 429-434 (1986), Coen et al, *Cell*, 47, 285-296 (1986) and Beld et al, *Plant Mol. Biol.*, 13, 491-502 (1989)). The other class comprises regulators of the biosynthetic genes (see Almeida et al, *Genes Dev.*, 3, 1758-1767 (1989), Harker et al *ibid*, Dooner and Nelson *Genetics*, 91, 309-315 (1979) and Beld et al *ibid*). This class includes the *C1* and *R* genes in maize, which encode products related to the *myb* and *myc* families of transcription factors respectively (see Paz-Ares et al, *EMBO J.*, 6, 3553-3558 (1987) and Ludwig et al, *Proc. Natl. Acad. Sci.*

U.S.A., 86, 7092-7096 (1989)).

The present invention is based on the isolation and characterisation of a gene designated *delila* that regulates pigmentation pattern in *Antirrhinum majus* and the use of this gene to regulate the expression of one or more anthocyanin pigment genes in a plant. Wild-type *A. majus* flowers have five red petals united to form a corolla tube with five distinct lobes. The epidermal cells of the petals contain red anthocyanin pigments. A recessive *delila (del)* mutation is known which confers a strikingly different pattern of floral pigmentation in which the corolla tubes are ivory and the lobes fully pigmented. The *del* mutation also blocks pigmentation of the anther filaments and lower stems and reduces that of the styles, sepals, carpels and petioles (leaf stalks). The wild-type *del* product is required in the corolla tube for normal transcript levels of many of the anthocyanin biosynthetic genes (see Almeida et al, *ibid* and Martin et al *ibid*). Although pigmentation of the corolla lobes is normally unaffected by *del*, in certain genetic backgrounds an effect of *del* in the lobes is revealed suggesting that *del* can also act in lobes.

As described in more detail below, the *del* locus of *A. majus* has now been cloned by a method involving transposon tagging and has been found to encode a potential protein *(DEL)* of 644 amino acids. The cDNA sequence of the cloned *del* locus and the deduced amino acid sequence of the *DEL* protein are shown in SEQ ID NO 1 and the deduced amino acid sequence of the *DEL* protein is shown in SEQ ID NO 2.

Summary of the Invention.

According to one aspect, the present invention provides a method for regulating the expression of one or more anthocyanin pigment genes in a plant which comprises the steps of transforming plant tissue with an expression vector comprising a DNA segment encoding a protein having

the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 and which is capable of regulating expression of one or more plant genes involved in pigment biosynthesis, the said DNA segment being under the control of a promoter upstream of and operably linked thereto and regenerating from the transformed tissue plants showing altered anthocyanin pigmentation. Preferably the DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 is a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2.

According to another aspect, the present invention provides a plant having a DNA segment as defined above incorporated into its genome or plant propagation material (such as seeds) of such a plant.

According to a still further aspect, the present invention provides a DNA molecule encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2 with the *DEL* protein as shown in SEQ ID NO 1 or 2 and which is capable of regulating expression of one or more plant genes involved in pigment biosynthesis.

According to a still further aspect the present invention provides the use of the DNA molecule encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO's 1 or 2 or a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2 or the protein encoded thereby to isolate a DNA molecule encoding a protein having the amino acid sequence which shows substantial homology

with the *DEL* protein as shown in SEQ ID NO 1 or 2 from other plant species.

According to a still further aspect the present invention provides an expression vector comprising a DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2 and which is capable of regulating expression of one or more plant genes involved in pigment biosynthesis, the said DNA segment being under the control of a promoter upstream of and operably linked thereto. The invention also provides a protein which is the product of expression of the expression vector as defined above in a host cell.

According to a still further aspect, the present invention provides a construct which comprises a transposon having cloned therein a DNA segment as defined above, the said DNA segment being under the control of a minimal promoter upstream of and operably linked thereto.

According to a still further aspect of the invention the present invention provides a method of trapping a promoter/enhancer which comprises the steps of introducing the construct into plant by transformation and propagating from said plant, plants having a phenotype showing altered anthocyanin pigmentation arising as a consequence of transposition of the construct.

According to a still further aspect of the invention the present invention provides a method for isolating a trapped promoter/enhancer from a plant which has been transformed with the construct as defined above which comprises reisolating the construct from said plant together with sequences adjacent thereto.

According to a still further aspect of the invention the present invention provides a method of expressing a gene of interest in a plant, which comprises transforming a cell of said plant with a first construct having said

gene of interest under the control of a first promoter, which first promoter is that of an anthocyanin gene, upstream of and operably linked thereto, the said plant having incorporated into its genome a DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 under the control of a second promoter upstream of and operably linked thereto, or the said plant being co-transformed with a second construct which comprises said DNA segment under the control of a third promoter, which third promoter may be the same or different to the second promoter, upstream of and operably linked thereto, or the said first construct optionally including the said DNA segment under the control of said second or third promoter upstream and operably linked thereto if the said plant does not have in its genome the said DNA segment or is not co-transformed with the said second construct, and deriving from the transformed plant further plants expressing said gene of interest.

According to a further aspect of the invention the present invention provides a method of expressing a gene of interest in a plant which comprises transforming said plant with a construct, which construct comprises a transposon having cloned therein a DNA segment as defined above, the said DNA segment being under the control of a minimal promoter upstream of and operably linked thereto, deriving from the transformed plant further plants having a phenotype showing altered anthocyanin pigmentation, reisolating from said plant the said construct together with sequences adjacent thereto, replacing said DNA segment in said construct with a gene of interest to form a new construct and transforming said plant with said new construct.

The cDNA encoding the *DEL* protein as shown in SEQ ID NO 1 or 2 contains a long open reading frame (ORF) starting at position +25. The ORF encodes a potential protein, *DEL*, of 644 amino acids which shows strong homology to the products of *Lc* and *R-S*, two members of the *R* gene family which controls pigmentation pattern in maize.

Maize and *Antirrhinum* are taxonomically distant and belong to the monocotyledoneae and the dicotyledoneae respectively, two groups thought to have diverged about 200 million years ago at an early stage in the evolution of flowering plants. There are marked differences in morphology and pigmentation pattern between the two species. The flowers of *Antirrhinum* are pollinated by bees and have large, vividly pigmented petals. In maize, which is wind-pollinated, the flowers are inconspicuous and there is no organ with obvious homology to petals. The organ most commonly pigmented is the seed, although the diverse alleles of the *R* gene family can pigment most plant tissues.

The structural and functional homology between the *DEL* protein and the proteins encoded by the *R* gene family of maize strongly suggests that the control of pigmentation pattern is mediated by a common regulator in different species, in spite of wide differences in morphology and coloration. Accordingly the present invention is not confined to the regulation of the expression of anthocyanin pigments in *Antirrhinum* but extends generally to all plants where a protein homologous to the *DEL* protein has an effect on the expression of anthocyanin pigments. The plant preferably belongs to the dicotyledoneae.

The manner in which the cDNA encoding the *del* locus of *Antirrhinum majus* has been cloned is described in more detail below. By making use of the cDNA sequence set out in SEQ ID NO 1, a DNA molecule encoding the *DEL* protein can now be obtained as required using standard techniques of cDNA cloning and/or DNA synthesis. DNA molecules encoding a protein having an amino acid sequence which shows

homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 can be obtained by mutation of a DNA molecule having the sequence shown in SEQ ID NO 1 using standard techniques of recombinant DNA technology and/or by DNA synthesis. Utilizing standard techniques the DNA molecule encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO's 1 or 2 or a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2 or protein encoded thereby may also be used to isolate DNA molecules encoding a protein having an amino acid sequence which shows homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 from other plant species, most preferably plant species belonging to the dicotyledoneae.

For use according to the present invention the DNA segment encoding the *DEL* protein or a protein homologous thereto will generally be incorporated in an expression vector which also includes suitable regulatory and control sequences to enable expression of the segment in a particular plant or part of a plant. Examples of suitable promoters include the cauliflower mosaic virus 35S promoter and also any promoter which is expressed in epidermal cells of different plant organs, such as the promoter of a housekeeping gene or a gene for the synthesis of specific epidermal structures. By use of a promoter which is specific for a particular type of plant tissue, i.e. is effective only in that tissue, the effect of the *DEL* protein or the protein homologous thereto can be confined to that specific tissue.

Plants transformed with the DNA segment encoding the *DEL* protein or a protein homologous thereto may be produced by standard techniques which are already known for the genetic manipulation of plants. For example the DNA segment may be incorporated into an *Agrobacterium* vector and plant material may then be infected by a strain of *Agrobacterium* carrying this vector. In this way the DNA

encoding the *DEL* protein or a protein homologous thereto becomes integrated into the genome of the plant tissue so that plants propagated from the tissue also carry this DNA. Alternative methods for the introduction to the DNA into plant cells include precipitation onto tungsten particles and shooting using a particle gun.

The ability to mediate or control the expression of anthocyanin pigmentation in plants can be put to practical use in a number of ways. Thus plant pigmentation can be increased or altered by transforming plants in the manner described above with a construct including the DNA encoding the *DEL* protein or a protein homologous thereto under control of a suitable promoter. By use of a construct in which a regulatory sequence, such as the promoter, is specific to a particular part of the plant, for example specific parts of the flower, the effect on pigmentation can be confined to that part of the plant.

In particular, the procedure described above can be used to enhance pigmentation of regions already pigmented in the host species or to pigment areas which are not normally pigmented in the host species. One specific application of this procedure is to produce novel genetically manipulated flowering plants with a phenotype in which the flowers show a pattern or intensity of pigmentation which differs from the host species.

It is possible to use the DNA segment encoding the *DEL* protein or a protein homologous thereto as a promoter/enhancer trap wherein the DNA segment which is driven by a "minimal" promoter, i.e. a truncated promoter more or less deficient in cis-acting regulatory sequences, may be cloned within a transposon (which can only transpose when a second gene containing the trans-activator is also present) and the construct is then introduced into plants. The transposon containing the DNA segment can then transpose to a new site near to a variety of promoters/ enhancers which can increase or activate transcription of the DNA segment. It is thus possible to derive and select

new pigmentation patterns by simply screening progeny from the transgenic plants. The transposon can subsequently be "stabilised" by crossing out the factor that activates the transposon. The chimeric sequence combining the DNA segment with the trapped promoter/enhancer can be recovered by reisolating back the construct, together with adjacent sequences, from the plant and used to control the expression of any gene of interest.

In particular the trapped promoter/enhancer may be used to control expression of a gene of interest in two ways, namely:

1) A new construct is prepared which comprises a gene of interest under the control of an anthocyanin gene promoter upstream and operably linked thereto. When this new construct is transformed into a plant, wherein a promoter/enhancer has been trapped, expression is seen of the gene of interest in those cells which express *delila*.

The trapped promoter/enhancer causes expression of the *DEL* protein which in turn switches on the anthocyanin promoter thus causing expression of the gene of interest.

2) In the re-isolated construct the DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 is excised from the construct and a heterologous gene of interest inserted in its place to form a new construct. When this new construct is transformed into a plant the heterologous gene of interest is expressed.

It may also be desirable to reduce plant pigmentation either in localised areas or throughout a plant. This may be achieved by various techniques such as those, for example, based on the use of DNA sequences showing sequence homology to *del* including antisense RNA, co-suppression or ribozymes. These techniques may be defined as follows:
Antisense RNA is where a gene is expressed in the opposite sense to normal (i.e. the promoter is at the 3′ end of the

gene), such that the "wrong" strand of the DNA is transcribed into RNA (giving antisense RNA). This antisense RNA may form a duplex with normal sense RNA and so inactivate it.

Co-suppression occurs where extra copies of a gene are introduced into the genome which may result in inactivation of the endogenous gene, which may in turn cause a mutant phenotype.

A Ribozyme is an RNA molecule which has the property that when it hybridizes to another RNA molecule containing a particular nucleotide sequence (target sequence), it will cleave the molecule and hence inactivate it. The nature of the target sequence depends on the critical region in the ribozyme molecule which is complementary to the target. The particular RNA molecule that the ribozyme recognises can therefore be altered by changing the critical region of the ribozyme.

In addition, constructs can be developed with alterations in the *del* coding sequence which produce proteins which interfere with the functioning of *delila* or *delila*-like genes in the host species.

The *del* coding sequence or a homologue thereof can also be used as a convenient visible marker. Thus use of a DNA sequence encoding the *DEL* protein or a protein homologous thereto in the manner described above allows the coding sequence in question to act as a visible marker for gene expression. This can be exploited to enable easy identification of transformed cells, cells in which a particular promoter is active, cells in which gene functions have been activated or inhibited, e.g. by excision or integration of a transposon. A DNA segment encoding the *DEL* protein or a protein homologous thereto in the manner described above may be used as a visible marker in a transgenic plant line, most particularly including dicotyledonous species, wherein a specific pigmentation pattern may be used to identify the line.

The *del* coding sequence or a homologue thereof can also be used to *trans*-activate or inhibit genes by placing the gene of interest under the control of a promoter, e.g. the *pallida* promoter of *Antirrhinum*, already known to be regulated by the *delila* gene. Plants containing such constructs, together with the *del* coding sequence (or an appropriate homologue thereof), would express the gene of interest only in those cells which express *delila*. This could be combined with the use of the *del* coding sequence as a visible marker so that cells expressing the gene of interest could be identified by their pigmentation phenotype.

A further possible use for the *del* coding sequence is in the isolation of homologous of the *delila* gene from various plant species. Such homologous of the *delila* gene can be isolated using genomic or cDNA probes derived from *delila* clones or based on the *del* coding sequence as set out in SEQ ID NO 1.

The invention is based on and further illustrated by the following experimental work in which reference is made to the accompanying drawings. It is to be noted that the experimental work utilizes *del*, however, clearly other DNA segments which encode a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 and which is capable of regulating expression of one or more plant genes involved in pigment biosynthesis. A brief description of each figure of the accompanying drawings is as follows.

FIGURE 1 - nucleotide and predicted amino acid sequence of *del* cDNA.

FIGURE 2 - Southern blots of *Eco*R1 - digested genomic DNA from various *Antirrhinum majus* plants.

FIGURE 3 - Sequence comparisons of Del^{+} and *del*-602 alleles in the region of the Tam2 insertion.

FIGURE 4 - Northern analysis of *del* expression in

various *Antirrhinum majus* flowers at different stages of development.

FIGURE 5 - Amino acid sequence comparison of *DEL* protein with selected HLH proteins.

FIGURE 6 - *In situ* hybridisation of medial longitudinal sections of corollas with ^{35}S labelled RNA probes.

FIGURE 7 - Plasmids pBJIMM15, pBJIMM21 and pBJIMM24.

1. Isolation of the *del* gene

A large scale transposon mutagenesis experiment yielded various homeotic and pigmentation mutants (Carpenter and Coen, *Genes Dev.*, 4, 1483-1493, (1990), Luo et al., *The Plant Journal*, 1, 59-69 (1991). One mutation gave a phenotype similar to that of the existing *del* mutation, and was shown to be a recessive *del* allele, *del*-602 (Luo et al., *ibid*). Self pollinated *del*-602 plants gave about 7.5% of progeny with a wild-type phenotype (revertants), explicable if a transposon excision occurred in germinal tissues of the parent plants. To identify the transposon at the *del* locus, genomic DNA from *del*-602 mutant and revertant plants was digested with restriction enzymes and probed with the various transposable elements isolated from *A. majus*. Because each of these elements was present in multiple copies in the genome, several bands were seen in Southern blots. When *Eco*RI digested DNA was probed with a fragment of the transposon Tam2 (Upadhaya et al., *Mol. Gen. Genet.*, 199, 201-207 (1985), a 5.6 kb band was consistently observed in mutants and not in revertants (Figure 2a). This suggested that the *del*-602 mutation resulted from a Tam2 insertion, and the 5.6 kb fragment was therefore cloned.

The resulting clone, pJAM 602, contained a 4.9 kb fragment of Tam2 with 0.7 kb of flanking DNA (Figure 2b). The flanking sequence (probe A) was then used to probe *Eco*RI digested DNA from various genotypes: *del*-602 plants showed the expected 5.6 kb band; plants of the progenitor

stock showed a wild-type band of 6.2 kb; and plants homozygous for a stable *del* allele, *del*-8, gave a 2.4 kb band (Figure 2b). If probe A derived from the *del* locus, reversion of the *del*-602 allele to wild-type should have correlated with restoration of the wild-type 6.2 kb band, and this was observed; six revertant plants obtained in the progeny of three crosses between *del*-602 and *del*-8 plants, and therefore representing at least three independent germinal reversions of the *del*-602 allele, all showed the 6.2 kb band, confirming that the pJAM 602 clone contained part of the *del* locus.

A clone of the *Del*$^+$ genomic region was obtained by screening a genomic library from the progenitor stock with probe A of pJAM 602 (Figure 2b). A comparison of the sequences flanking the Tam2 insertion in the *del*-602 allele with the corresponding wild-type sequences identified a direct duplication of 3 base pairs of target DNA, a length characteristic of Tam2 insertions (Upadhaya et al, *ibid*) (Figure 3).

Figure 2(a) shows Southern blot of *Eco*RI-digested genomic DNA from *del*-602 mutant and revertant (*Del*$^+$) plants, probed with a 4.4 kb *Eco*RI/*Hind*III fragment of the Tam2 clone pRH2, provided by Enno Krebbers and Hans Sommer. These plants were obtained in the F_1 progeny of a cross between *del*-602 and *del*-8 plants. Revertants have the presumed genotype *Del*$^+$/*del*-8, and mutants *del*-602/*del*-8. A restriction map of Tam2 and the origin of the probe are shown below the autoradiograph. Sites indicated are *Eco*RI (E), *Hind*III (H) and *Bgl*II (B).

Figure 2(b) shows Southern blot of *Eco*RI-digested genomic DNAs probed with fragment A of the 5.6 kb *Eco*RI clone, pJAM 602. Lane 1, wild-type progenitor of *del*-602; lane 2, homozygous *del*-602; lane 3, homozygous *del*-8; lanes 4-9, revertant progeny from crosses between *del*-602 and *del*-8 plants; lanes 10-16 *del* mutant progeny from the same crosses. A restriction map of the 5.6 kb *Eco*RI clone, pJAM 602, and the origin of the probe are shown below. Thick

line, Tam2 sequences; thin line, flanking sequences.

RNA extraction and Southern blot analysis were performed as described by Coen et al., *Cell*, 47, 285-296 (1986). The pJAM 602 clone was obtained by digesting genomic DNA of *del*-602 plants with *Eco*RI, gel-purifying fragments in the 5-6 kb size range, ligating to λNM1149 arms and screening a library of 30,000 plaques with the Tam2 probe shown in Figure 2a. The resulting clone was subcloned into Bluescript SK$^+$ (Stratagene).

Figure 3 shows a sequence comparison of *Del*$^+$ and *del*-602 alleles in the region of the Tam2 insertion. The target sequence, duplicated on insertion of Tam2, is boxed. The pJAM 602 clone was sequenced to provide flanking sequences to the right of the Tam2 insertion. To obtain the flanking sequence to the left of Tam2, 0.1 μg of genomic DNA from a *del*-602 plant was amplified by PCR (Saiki et al, *Science*, 239, 487-491 (1988), using a primer derived from sequences near the left terminus of Tam 2 (as orientated in Figure 2b) and a second primer based on *Del*$^+$ genomic sequences. The expected fragment of 0.3 kb was subcloned to give pJAM 122 and sequenced.

2. Characteristics of the putative *DEL* protein and expression of *del*

To identify the *Del*$^+$ transcript, polyA$^+$ RNA extracted from corolla tubes was hybridised with probe A of pJAM 602. A single 2.5 kb transcript, detected in wild type tubes, was absent from the flowers of both *del*-602 and *del*-8 mutants (Figure 4a). A cDNA library prepared from wild type flower buds was screened with probe A and several cDNA clones were isolated and sequenced. The cDNA sequence (SEQ ID NO 1) contained a long open reading frame (ORF) starting at position +25 with an ATG codon flanked by sequences which conformed to the consensus for initiation of translation in plants (Lüttke et al., *EMBO J.*, 6, 43-48 (1987). The ORF encoded a potential protein, *DEL*, of 644 amino acids. Comparison of the amino acid sequence of *DEL* to proteins on the PIR and SWISS databases using the FASTA

program (Lipman and Pearson, Science, 227, 1435-1441 (1985) revealed a strong homology between *DEL* and the products of *Lc* (Ludwig et al., *ibid*) and *R-S* (Perrot and Cone, *Nucl. Acids Res.*, 17, 8003 (1989)), two members of the *R* gene family which control the pigmentation pattern in maize. The deduced amino acid sequences of the products of *del*, *R* and *Lc* had 38% identity when optimally aligned (58% similarity, allowing for conservative changes). Two regions of *DEL* were highly conserved; near the N-terminus, residues 16-190 were 61% identical with the corresponding regions of *R-S* and *Lc* proteins, and towards the carboxy terminus, residues 438-497 were 60% identical. The latter region also resembled a helix-loop-helix (HLH) domain conserved in a number of eukaryotic regulatory genes (SEQ ID NO's 3 to 12). The strongest similarity was to *myc* proteins, involved in animal cell proliferation control, and the human transcription factor E3, which binds to the immunoglobin heavy chain enhancer motif μE3.

DEL also contained a highly acidic region (residues 173-319); 27 acidic and two basic residues gave an overall negative charge of -25. A corresponding acidic region occurs in *Lc* (Ludwig et al., *ibid*), but relatively little conservation in amino acid sequence was found between these regions of *Lc* and *DEL* (25% identity). The conserved region near the N-terminus of *DEL* showed no significant homology to proteins other than those of the *R* gene family. Secondary structure analysis (Chou and Fasman, *Biochemistry*, 13, 211-244 (1974)) of this region identified several sequences predicted to form α helices, one of which (residues 79-89) appeared strongly amphipathic in helical wheel plots and was highly conserved between *DEL* and *Lc*.

Alignment of the sequence of the *del*-602 genomic clone with the cDNA sequence showed that Tam2 was inserted in an intron, 285 nucleotides upstream of the 3' intron-exon junction (Figure 1 and SEQ ID NO 1). Northern analysis of RNA from *del*-602 flowers revealed various aberrant-sized transcripts (Figure 4a), which could result from the Tam2

insertion interfering with processing of the primary *del* transcript.

A *del* cDNA clone was used to screen a wild-type genomic library and three clones with extensive homology to *del* were isolated. Detailed restriction mapping indicated that they derived from independent loci, distinct from *del*. Therefore, A *majus* has a family of at least four genes related to *del*.

To determine the temporal pattern of *del* expression, a Northern blot of RNA from different-sized flower buds was probed with a *del* cDNA (Figure 4b). A low level of expression was detected in the youngest flower buds examined and was maintained until shortly before anthesis (correlated with flower opening) when expression increased and then declined. The rise in *del* expression occurred at about the same time that flowers became strongly pigmented. However, there was detectable expression considerably before flower buds were visibly pigmented. This could imply that a threshold level of *del* product was necessary for anthocyanin biosynthesis, or that additional factors were required at these early stages for *del* to activate its target genes. To compare the expression of *del* with one of its target genes, the Northern blot was stripped and reprobed with a cDNA clone of *pallida* (*pal*), a biosynthetic gene strongly regulated by *del* (Figure 4b. Almeida et al, *ibid*). At the later stages of flower development the pattern of *pal* expression closely resembled that of *del*, but the earliest detectable *pal* expression was after that of *del*, implying that *del* expression was not sufficient to activate *pal* at very early stages. The spatial pattern of *del* expression in the corolla was determined by Northern analysis of RNA from dissected wild-type corollas (Figure 4c). Pigmentation of the corolla first appeared in the lobes and in a ring at the base of the tubes, and subsequently extended throughout the tubes. The strongest expression of *del* was seen at the base of the tubes, the region of pigmentation most greatly affected by the *del*

mutation. Expression was also detected in the lobes, as was predicted from genetic interactions described previously (Almeida et al., *ibid*). As expected, no wild-type *del* transcript was found in the pigmented lobes of *del* mutant flowers (results not shown), confirming that *del* expression is not required for pigment biosynthesis in the lobes. Probing of a Northern blot of RNA from diverse organs with *del* cDNA showed the strongest expression in the corolla, stamens and style, with a weak signal in sepals and carpels, and little or none in bracts, leaves and stem (Figure 4d). All organs in which *del* expression was detected were visibly pigmented, and the level of *del* expression correlated with the degree of pigmentation.

The expression of *del* was further localised by *in situ* hybridisation of ^{35}S-labelled *del* RNA to sections of wild-type corollas. Signal was detected only when the antisense strand of *del* was used as a probe, and was strongest in the flower buds 1-6 nodes above the first fully opened flower on the inflorescence. The signal was specific to the epidermal cell layers in both tubes and lobes (Figure 6). This corresponds to the distribution of anthocyanins, and of expression of the biosynthetic genes *nivea*, *pallida* and *incolorata* (Jackson, *Current Biology*, 1, 99 (1991)). The two epidermal layers of the petal are referred to as inner (lining the throat of the tube and contiguous surface of the lobes) and outer (lining the exterior surface of the tube and lobes). In the corolla tubes, signal was as intense in the outer epidermis as in the inner epidermis (Figure 6a). The outer epidermis of the tubes and lobes bore numerous multi-cellular hairs, which were unpigmented; no signal was seen in these hairs (data not shown). In the corolla lobes, a strong signal was also seen in both the outer and inner epidermi. However, in sections from the central face of the flower, the signal was most abundant in the inner epidermis (Figure 6b), which also had strongest pigmentation.

Figure 4 shows Northern analysis of *del* expression.

(a) Northern blot of RNA from the corolla tubes of *DEL*$^+$, *del*-8 and *del*-602 flowers, hybridised with probe A of pJAM 602 (Figure 2b). Each lane contained 3μg of polyA$^+$RNA. Note the hybridizing band at 2.5 kb. (b) RNA from flower buds of different sizes, as indicated by diagrams above each track. Nodes on the *Antirrhinum* inflorescence (a raceme) bear flower buds at progressively older developmental stages in a series from apex to base. RNA was extracted from flowers at different positions on the inflorescence; 21-28 nodes above the first fully opened flower (lane 1), 17-20 nodes above (lane 2), 13-16 nodes above (lane 3), 9-12 nodes above (lane 4), 5-8 nodes above (lane 5), 1-4 nodes above (lane 6), and the first four fully opened flowers (lane 7).

The Northern blot was first probed with the *del* cDNA clone pJAM 121, then after autoradiography it was stripped and reprobed with a *pallida* (*pal*) cDNA clone pJAM 225, provided by C. Martin. The *pal* gene encodes the enzyme dihydroflavonol reductase involved in anthocyanin biosynthesis, and is known to be regulated by *del* (Almeida et al, *Genes Dev.*, *3*, *1758-1767 (1989)*). *(c)* RNA from wild-type flower buds (1-5 nodes above first fully open flower) dissected into three parts: the base of corolla tubes (BT), the rest of the tube (RT), and the lobes (L) as indicated above the autoradiogram, probed with the *del* cDNA clone pJAM 121. (d) RNA from leaves (L), stems (S), bracts (B), sepals (Se), petals (P), stamens (St), styles (Sty) and carpels (Ca), probed with the *del* cDNA clone pJAM 121. Each lane of (b), (c) and (d) contains 10μg of RNA, and loading appeared equal when ribosomal bands were viewed under UV illumination after staining with ethidium bromide.

RNA extraction and Northern analysis was carried out as described (in Coen et al., *ibid*). Northern blots were stripped by washing in 0.5% SDS, 0.01 % SSC for 30 minutes at 80°C.

Nucleotide and predicted amino acid sequence of *del* cDNAs is shown in Figure 1 and SEQ ID NO 1. The predicted

amino acid sequence of *del* is also shown in SEQ ID NO 2. The solid triangle indicates the position of an intron within which Tam2 is inserted in the *del*-602 allele, namely between bases 573 and 574. The region of the *DEL* protein with similarity to the helix-loop-helix domain of the *myc* family of transcription factors is underlined, it commences at residue 439 and terminates at residue 493. Residues conserved in *DEL* and the maize R-S gene product (after alignment) are shaded.

Figure 5 and SEQ ID NO's 3 to 12 show amino acid sequence comparison of *DEL* protein with selected HLH proteins. Alignments were made to maximise homology within the HLH domain. A consensus sequence derived for residues conserved in most known HLH genes is shown below (Benezra et al., *Cell*, 61, 49-59 (1990) and Cai and Davis, *Cell* 61 437-446 (1990)) (Ψ = L,I,V,M). Shaded regions identify residues that match the consensus. The positions of the conserved basic region, putative amphipathic helices I and II, and the loop are shown above (Murre et al, *Cell*, 56, 577-783 (1989)). The sequences shown are for: maize *R-S* (Perrot and Cone, *ibid*) and *Lc* (Ludwig et al., *ibid*); human *E3* (Beckman et al, *Genes Dev.*, 4, 167-179 (1990)), *L-myc* (DePinho et al, *Genes Dev.*, 1, 1311-1326 (1987) and *N-myc* (Kohl et al, *Nature*, 319, 73-77 (1986)); yeast *Cbf1* (Cai and Davis, *ibid*); AP4 (Hu et al, Genes Dev, 4, 1741-1752, (1990)); B (Radicella et al, Plant Mol. Biol., 17, 127-130 (1991)) and mouse *myogenin* (Edmondsen and Olson, *Genes Dev.*, 3, 628-640 (1989)).

cDNA was synthesised from 3μgpolyA$^+$RNA extracted from wild-type flower buds (1-4 nodes above first fully opened flower), and cloned into the *Eco*RI site of λNM1149 using Amersham kits. A library of 10^5 plaques was screened with probe A of pJAM 602 (Figure 2b) and the longest clone obtained was subcloned into Bluescript SK$^+$ (Stratagene). Sequence analysis revealed that the cDNA insert lacked a poly A tail but contained a long open reading frame which terminated at an *Eco*RI site, suggesting that the cDNA had

been cleaved at an internal site during the cloning. Alignment of genomic sequences of pJAM 602 with the cDNA showed that the probe terminated at the corresponding *Eco*RI site, so the library was rescreened with a fragment of a genomic *Del*$^+$ clone extending 3′ of the original probe. A further clone was obtained and contained an insert with sequences extending from the *Eco*RI site to the poly A tail. To confirm that the two sequences were contiguous, an intact clone, pJAM 121, was isolated by PCR amplification of cDNA ends (3′ RACE) (Frohman et al., Proc. Natl. Acad. Sci. U.S.A., 85, 8998-9002 (1988)) using a specific primer based on sequences from the 5′ end of the cDNA sequence, and this was sequenced around the *Eco*RI site. The *del* cDNA sequence is about 0.4 kb smaller than the observed transcript, suggesting that it is not full length. However, alignment of *DEL* with *Lc*/R proteins suggests that the cDNA sequence contains the entire *del* coding sequence. Sequences were determined by the plasmid dideoxynucleotide sequencing method (Chen and Seeburg, DNA 4, 165-170 (1985)) using a Sequenase kit (USB), and both strands of cDNAs were sequenced. Computer analysis of sequences was performed using the University of Wisconsin Computer Group programs.

Figure 6 shows *in situ* hybridisation of medial longitudinal sections of corollas with ^{35}S labelled RNA probes. (a) Tube tissue under light (left) and dark field (right). The inner epidermis (I), mesophyll (M) and outer epidermis (O) are labelled. Silver grains corresponding to *del* expression are seen in dark field, and are localised to the epidermis in inner and outer surfaces of the tube. (b) Lobe tissues viewed under light field (left) and dark field (right). Silver grains are at high density over the inner epidermis, and at a lesser density over the outer epidermis.

Corollas were fixed in 4% paraformaldehyde and embedded in wax, and sections were prepared for hybridisation as described by Jackson, *Molecular Plant Pathology: A Practical Approach (Bowles, Gurr, McPherson*

eds) Oxford University Press (1991). A 1kb fragment of the *del* cDNA encoding the N-termimus but not the HLH region of *DEL* was subcloned into Bluescript SK^+ and KS^+. These plasmids were linearised with restriction enzymes cutting in the polylinker furthest from the T7 promoter. ^{35}S RNA probes were synthesised using 1μg of linearised plasmid as template, t7 polymerase and 50 μCi of ^{35}S UTP (1300 mCi/mmol, NEN). Transcription reactions, hybridisation, and autoradiography were carried out as described by (Ingham et al, *Nature*, 318 439-445 (1985)). Autoradiographic exposure was for 14 days at 4°C. No signal above background was observed using control sense-strand probes.

3. Transformation of both tomato and tobacco with the *del* gene

A number of independent transformed lines, of both tomato and tobacco, have been produced using three molecular constructs that utilise the *del* gene. The transformation of both tomato and tobacco was performed through a standard procedure utilising *Agrobacterium tumefaciens* as detailed below.

Tomato: Approximately 100 tomato seed (*Lycopersicon esculentum* variety Money Maker) were surface sterilised using a 10% aqueous solution of bleach and sown on agar media under sterile conditions (day 0). The seeds were allowed to germinate and grow for 10 days (day 10).

On day 9 a 10ml volume of luria broth was inoculated with *Agrobacterium tumefaciens* strain Ach5 carrying the plasmid pAL4404 (Hoekema et al, *Nature*, 303, 179-180 (1983)) and one of the following plasmids, pBJIMM15, pBJIMM21 or pBJIMM24. The plasmids can be seen in Figure 7 and are derived from the plasmid pSLJ456 (derived by J. D. Jones & C. Dean et al from pRK290 J. D. Jones & C. Dean et al P.N.A.S., 77, 7347, (1980)). As can be seen in the diagrams pBJIMM15 carries the cDNA of the *del* gene of *Antirrhinum majus* under control of the cauliflower mosaic virus 35S promoter and is terminated by the OCS gene 3'

terminator sequence, pBJIMM21 carries a copy of the *del* gene derived from a genomic fragment again driven by the 35S promoter, and pBJIMM24 carries a large genomic fragment containing both the promoter and coding sequence of *del*. These three plasmids all also carry the NPT gene that confers kanamycin resistance and this gene is driven by the 2'1' promoter (Velten et al. EMBO J., 3, 2723-2730, (1984)) again terminated by the OCS 3' sequence. The three different cultures (differing only in the type of pBJIMM plasmid that they contained) were allowed to grow for two days (until day 11). The transformation procedure employed was identical for each type of culture and therefore will only be detailed once.

On day 11 the tomato seedlings were cut into small pieces to produce tissue explants. These explants were washed with the culture of *Agrobacterium tumefaciens* described above and the bacterium was allowed to remain in contact with the explants for 2 days (to day 13).

On day 13 the tissue explants were placed on selective agar plates bearing appropriate antibiotics to kill any remaining *Agrobacterium* and allow only "transformed" plant tissue to regenerate. The period between days 11 and 13 when the *Agrobacterium* is co-cultivated with the tomato tissue is to allow time for the *Agrobacterium* to transfer the portion of DNA from the pBJIMM plasmids that lie between the points marked "left border" and "right border" in Figure 7 into the genome of the tomato. Once incorporated into the plant genome, the NPT gene becomes functional confers kanamycin resistance to the tomato tissue and thereby allowing it to regenerate on agar medium containing kanamycin.

Over the next two to six weeks (days 27 to 55) calli and shoots regenerated from the explants. Small shoots were removed and transplanted into media containing kanamycin. Only shoots that derive from "transformed" tissue will produce roots on this media. For all pBJIMM constructs rooting was observed with on average 10% of

shoots transferred to Kanamycin containing rooting media. Rooting was found to occur over one to three weeks (day 62 to 76). When rooted shoots were approximately 5cm tall they were transferred to peat based potting compost and grown on in the glasshouses for 6 weeks (to day 118).

Through this procedure 10 plants were regenerated from the pBJIMM15 transformation seven of which have shown enhanced anthocyanin pigmentation phenotypes. 9 Plants from the pBJIMM21 transformation seven of which show enhanced pigmentation and one plant from the pBJIMM24 transformation that shows a small increase in pigmentation were obtained. Southern blot hybridisation analysis revealed that all plants showing pigmentation had incorporated the *del* construct from between the left and right border points of the pBJIMM plasmid. The phenotypes of the plants from the pBJIMM15 and pBJIMM21 transformations were very similar. The shoots and leaves were much more heavily pigmented with anthocyanin than the control plants. When flowers were produced they were observed to have a stripe of purple pigmentation visible over the main vein of the petal, a phenotype absence from the control. When the roots of these plants were exposed to light it was found that purple anthocyanin pigment was accumulated, in contrast to a very low level of pigmentation observed in controls. When these tissues were sectioned it was found that the pigment was being produced in the epidermal and sub-epidermal layers of the leaf, the epidermal and sub-epidermal layers of the stem and only in the sub-epidermal layer of the root.

Tobacco: The tobacco transformation procedure is very similar to that of tomato. The same cultures of *Agrobacterium* were used, however, plants have only been regenerated from pBJIMM15 and pBJIMM21 cultures. On day 0 10 ml volumes of Luria broth were inoculated with *Agrobacterium* cultures as previously described, in this case only cultures bearing pBJIMM15 and 24 have been used. On day 2 several immature leaves of tobacco (*Nicotiana*

tabaccum variety Samson) were harvested from plants growing in the greenhouse. These were then surface sterilized by washing in a 10% solution of bleach for five minutes and then rinsed in sterile water and cut into small tissue explants. The explants were then washed with the *Agrobacterium* cultures and allowed to co-cultivate for two days (days 3 to 4).

The explants were then transferred to agar plates containing kanamycin and over the next 2 to 3 weeks tissue regeneration occurred resulting in the production of small shoots (days 15 to 21). These were excised and rooted in media containing kanamycin, after producing roots and reaching a height of approximately 5cm these plantlets were transferred to a peat based potting compost and grown on in the greenhouse to maturity in approximately 2 months. Through this method 7 plants have been produced from the pBJIMM15 culture and of these 5 show enhanced pigmentation, 9 plants have been produced from the pBJIMM24 transformation and 4 show a slight enhancement of pigmentation. By southern analysis it was shown that plants which exhibited a change in pigmentation also carried the *del* construct from the pBJIMM plasmid. The pigment phenotype of the transformed tobacco was visible only in the flowers of the plant. The flower petals and anther filaments of the transformants were much more intensely pigmented than similarly cultivated control plants. On sectioning this pigmentation appeared to be epidermal in the flower and both epidermal and sub-epidermal in the anther filament. The phenotype was only observed in the flower as the shoots, stem and leaves of the transformants were all indistinguishable from the control plant.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: COEN, Enrico
(B) STREET: John Innes Institute, Colney Lane
(C) CITY: Norwich
(D) STATE: Norfolk
(E) COUNTRY: Great Britain
(F) POSTAL CODE (ZIP): NR4 7UH
(G) TELEPHONE: (0603) 52571
(H) TELEFAX: (0603) 56844
(I) TELEX: 975122 (JIINOR G)

(A) NAME: CARPENTER, Rosemary
(B) STREET: John Innes Institute, Colney Lane
(C) CITY: Norwich
(D) STATE: Norfolk
(E) COUNTRY: Great Britain
(F) POSTAL CODE (ZIP): NR4 7UH
(G) TELEPHONE: (0603) 52571
(H) TELEFAX: (0603) 56844
(I) TELEX: 957122 (JIINOR G)

(A) NAME: GOODRICH, Justin
(B) STREET: John Innes Institute, Colney Lane
(C) CITY: Norwich
(D) STATE: Norfolk
(E) COUNTRY: Great Britain
(F) POSTAL CODE (ZIP): NR4 7UH
(G) TELEPHONE: (0603) 52571
(H) TELEFAX: (0603) 56844
(I) TELEX: 957122 (JIINOR G)

(A) NAME: FREEDMAN, Roger
(B) STREET: Diatech Limited, 19 Millers Yard, 10/11 Mill Lane,
(C) CITY: Cambridge
(E) COUNTRY: Great Britain
(F) POSTAL CODE (ZIP): CB2 1RQ
(G) TELEPHONE: (0223) 69936
(H) TELEFAX: (0223) 464113

(ii) TITLE OF INVENTION: Regulation of Plant Genes

(iii) NUMBER OF SEQUENCES: 12

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(v) CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/818,570
(B) FILING DATE: 09-JAN-1992

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2075 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Antirrhinum Majus

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 25..1956

(ix) FEATURE:
(A) NAME/KEY: insertion_seq
(B) LOCATION: (573^574)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GTAGAGAGGA GAGAGGATTC AAGA ATG GCT ACT GGT ATC CAA AAC CAA AAG          51
                           Met Ala Thr Gly Ile Gln Asn Gln Lys
                             1               5

ATA GTC CCT GAG AAT TTG AGG AAG CAA CTT GCT ATT GCT GTG AGA AGT          99
Ile Val Pro Glu Asn Leu Arg Lys Gln Leu Ala Ile Ala Val Arg Ser
 10                  15                  20                  25

ATC CAA TGG AGT TAT GCA ATT TTC TGG TCC AAT TCA GTT GCA CAA CCA         147
Ile Gln Trp Ser Tyr Ala Ile Phe Trp Ser Asn Ser Val Ala Gln Pro
                 30                  35                  40

GGG GTC TTG GAG TGG GGT GAT GGG TTC TAC AAT GGA GAT ATT AAA ACT         195
Gly Val Leu Glu Trp Gly Asp Gly Phe Tyr Asn Gly Asp Ile Lys Thr
             45                  50                  55

CGA AAA ACT GTA CAA TCT GTC GAA TTG AAT CAA GAT CAG CTG GGA TTG         243
Arg Lys Thr Val Gln Ser Val Glu Leu Asn Gln Asp Gln Leu Gly Leu
         60                  65                  70

CAG AGA AGT GAT CAA TTG AGA GAA CTT TAT GAG TCT CTT TCA CTT GGT         291
Gln Arg Ser Asp Gln Leu Arg Glu Leu Tyr Glu Ser Leu Ser Leu Gly
     75                  80                  85
```

```
GAA ACC AAC ACA CAA GCT AAA AGG CCT ACT GCT GCA TTA TCA CCA GAA          339
Glu Thr Asn Thr Gln Ala Lys Arg Pro Thr Ala Ala Leu Ser Pro Glu
 90                  95                  100                 105

GAC CTC ACT GAT GCT GAG TGG TTT TTC TTG GTT TGC ATG TCT TTC ATA          387
Asp Leu Thr Asp Ala Glu Trp Phe Phe Leu Val Cys Met Ser Phe Ile
                110                 115                 120

TTC AAT ATT GGC CAA GGG TTG CCT GGA AGA ACA TTA GCA CGA AAT CAA          435
Phe Asn Ile Gly Gln Gly Leu Pro Gly Arg Thr Leu Ala Arg Asn Gln
            125                 130                 135

GCA GTA TGG CTA TGC AAC GCT CAT CGT GCG GAC ACC AAA GTT TTC TCG          483
Ala Val Trp Leu Cys Asn Ala His Arg Ala Asp Thr Lys Val Phe Ser
        140                 145                 150

CGT TCT TTG CTT GCA AAG AGT GCG TCA ATT CAG ACA GTT GTG TGC TTT          531
Arg Ser Leu Leu Ala Lys Ser Ala Ser Ile Gln Thr Val Val Cys Phe
    155                 160                 165

CCA TAT TCA GAA GGT GTA GTT GAG CTG GGA GCA ACA GAG CTA GTA CCG          579
Pro Tyr Ser Glu Gly Val Val Glu Leu Gly Ala Thr Glu Leu Val Pro
170                 175                 180                 185

GAG GAT TTG AAT CTA ATC CAG CAT ATA AAA ACT TCA TTC TTG GAC AGT          627
Glu Asp Leu Asn Leu Ile Gln His Ile Lys Thr Ser Phe Leu Asp Ser
                190                 195                 200

CCT GCC ACC GTT CCC AAG ATT CCC AAC TAT GTC TCC AAC AGT ATT ACA          675
Pro Ala Thr Val Pro Lys Ile Pro Asn Tyr Val Ser Asn Ser Ile Thr
            205                 210                 215

AAC AAC AAT GAC CTC ATT TGT GAA GCG CTT GAA CAT GCT AAT ATA CCA          723
Asn Asn Asn Asp Leu Ile Cys Glu Ala Leu Glu His Ala Asn Ile Pro
        220                 225                 230

GAA AAC GAT CTT GAT CAG CTT TTG AAT TGT CCA GAC ACG AAC ATA TGT          771
Glu Asn Asp Leu Asp Gln Leu Leu Asn Cys Pro Asp Thr Asn Ile Cys
    235                 240                 245

TCT CCT GAT AAC AGT TTG GAT GAC TTT GCA GAC AAT TTA CTC ATA GAC          819
Ser Pro Asp Asn Ser Leu Asp Asp Phe Ala Asp Asn Leu Leu Ile Asp
250                 255                 260                 265

GAA TCG AAT TTG GCA GAA GGC ATC AAT GGG GAG GTT CCT CAA ACA CAA          867
Glu Ser Asn Leu Ala Glu Gly Ile Asn Gly Glu Val Pro Gln Thr Gln
                270                 275                 280

AGC TGG CCT TTC ATG GAT GAT GCA ATC AGC AAT TGT CTC AAT AGT TCT          915
Ser Trp Pro Phe Met Asp Asp Ala Ile Ser Asn Cys Leu Asn Ser Ser
            285                 290                 295

ATG AAT TCT AGT GAC TGT ATA TCT CAA ACT CAT GAA AAT CTA GAG TCT          963
Met Asn Ser Ser Asp Cys Ile Ser Gln Thr His Glu Asn Leu Glu Ser
        300                 305                 310
```

```
TTT GCT CCA CTT TCT GAT GGA AAA GGG CCA CCG GAG ACG AAT AAT TGT      1011
Phe Ala Pro Leu Ser Asp Gly Lys Gly Pro Pro Glu Thr Asn Asn Cys
    315                 320                 325

ATG CAC AGC ACT CAA AAA TGC AAT CAG CAG ATA GAA AAC ACG GGT GTC      1059
Met His Ser Thr Gln Lys Cys Asn Gln Gln Ile Glu Asn Thr Gly Val
330                 335                 340                 345

CAA GGC GAT GAG GTC CAT TAT CAA GGG GTA CTT TCC AAT CTT TTG AAG      1107
Gln Gly Asp Glu Val His Tyr Gln Gly Val Leu Ser Asn Leu Leu Lys
                350                 355                 360

AGT TCC CAT CAG TTG GTT CTT GGT CCC TAC TTC AGA AAT GGG AAT AGA      1155
Ser Ser His Gln Leu Val Leu Gly Pro Tyr Phe Arg Asn Gly Asn Arg
            365                 370                 375

GAA TCA AGC TTC GTT AGT TGG AAC AAG GAT GGA TCG TCG GGT ACT CAT      1203
Glu Ser Ser Phe Val Ser Trp Asn Lys Asp Gly Ser Ser Gly Thr His
        380                 385                 390

GTT CCC CGA AGC GGA ACC TCA CAA AGA TTT CTG AAG AAA GTA CTT TTT      1251
Val Pro Arg Ser Gly Thr Ser Gln Arg Phe Leu Lys Lys Val Leu Phe
    395                 400                 405

GAA GTA GCT AGA ATG CAT GAA AAC TCC AGG CTT GAT GCT GGT AAA CAA      1299
Glu Val Ala Arg Met His Glu Asn Ser Arg Leu Asp Ala Gly Lys Gln
410                 415                 420                 425

AAG GGC AAC AGT GAC TGC CTT GCA AAG CCA ACG GCT GAT GAA ATT GAT      1347
Lys Gly Asn Ser Asp Cys Leu Ala Lys Pro Thr Ala Asp Glu Ile Asp
                430                 435                 440

AGA AAC CAC GTC TTG TCA GAG AGA AAA CGC AGA GAG AAA ATA AAC GAA      1395
Arg Asn His Val Leu Ser Glu Arg Lys Arg Arg Glu Lys Ile Asn Glu
            445                 450                 455

CGG TTT ATG ATT CTT GCA TCC CTA GTC CCA TCC GGT GGC AAG GTT GAC      1443
Arg Phe Met Ile Leu Ala Ser Leu Val Pro Ser Gly Gly Lys Val Asp
        460                 465                 470

AAA GTA TCA ATA CTA GAC CAT ACA ATA GAT TAC TTG AGA GGG CTT GAG      1491
Lys Val Ser Ile Leu Asp His Thr Ile Asp Tyr Leu Arg Gly Leu Glu
    475                 480                 485

AGG AAA GTC GAC GAG CTG GAA TCT AAC AAA ATG GTA AAG GGC CGG GGG      1539
Arg Lys Val Asp Glu Leu Glu Ser Asn Lys Met Val Lys Gly Arg Gly
490                 495                 500                 505

CGG GAA TCA ACT ACA AAA ACT AAA CTA CAC GAT GCC ATT GAG AGG ACC      1587
Arg Glu Ser Thr Thr Lys Thr Lys Leu His Asp Ala Ile Glu Arg Thr
                510                 515                 520

TCT GAT AAT TAT GGC GCA ACA AGG ACA ACT AAC GTC AAG AAA CCG TTG      1635
Ser Asp Asn Tyr Gly Ala Thr Arg Thr Ser Asn Val Lys Lys Pro Leu
            525                 530                 535
```

```
ACA AAC AAG AGA AAG GCT TCT GAT ACG GAC AAC ATT GGA GCC GTA AAT      1683
Thr Asn Lys Arg Lys Ala Ser Asp Thr Asp Lys Ile Gly Ala Val Asn
        540                 545                 550

AGC AGA GGT CGA TTG AAA GAT TCC TTA ACA GAT AAT ATA ACT GTG AAC      1731
Ser Arg Gly Arg Leu Lys Asp Ser Leu Thr Asp Asn Ile Thr Val Asn
    555                 560                 565

ATT ACA AAC AAG GAT GTG TTG ATT GTC GTG ACT TGT TCT TCC AAG GAG      1779
Ile Thr Asn Lys Asp Val Leu Ile Val Val Thr Cys Ser Ser Lys Glu
570                 575                 580                 585

TTT GTA TTG CTT GAA GTG ATG GAA GCC GTA AGA CGA CTA AGT TTG GAT      1827
Phe Val Leu Leu Glu Val Met Glu Ala Val Arg Arg Leu Ser Leu Asp
                590                 595                 600

TCC GAA ACT GTT CAA TCT TCC AAC AGA GAT GGA ATG ATA TCT ATT ACC      1875
Ser Glu Thr Val Gln Ser Ser Asn Arg Asp Gly Met Ile Ser Ile Thr
            605                 610                 615

ATA AAA GCC AAG TGC AAG GGA TTG AAG GTT GCA TCA GCA AGT GTG ATC      1923
Ile Lys Ala Lys Cys Lys Gly Leu Lys Val Ala Ser Ala Ser Val Ile
        620                 625                 630

AAA CAA GCT CTT CAG AAA GTT ACT ATG AAG TCT TGAAGTTGAT TTATGCTCAC    1976
Lys Gln Ala Leu Gln Lys Val Thr Met Lys Ser
    635                 640

TATCTATAGC TAGCTTTTGT GTAAAAAATT TGTATTCATA ACTTTTGCTA AGTAATTTGC    2036

AGGGCTTTTC CAAGTAGTTC AGATCAATAA AAAAAAAAA                           2075
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 644 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Thr Gly Ile Gln Asn Gln Lys Ile Val Pro Glu Asn Leu Arg
1 5 10 15

Lys Gln Leu Ala Ile Ala Val Arg Ser Ile Gln Trp Ser Tyr Ala Ile
20 25 30

Phe Trp Ser Asn Ser Val Ala Gln Pro Gly Val Leu Glu Trp Gly Asp
35 40 45

Gly Phe Tyr Asn Gly Asp Ile Lys Thr Arg Lys Thr Val Gln Ser Val
50 55 60

Glu Leu Asn Gln Asp Gln Leu Gly Leu Gln Arg Ser Asp Gln Leu Arg
65 70 75 80

Glu Leu Tyr Glu Ser Leu Ser Leu Gly Glu Thr Asn Thr Gln Ala Lys
85 90 95

Arg Pro Thr Ala Ala Leu Ser Pro Glu Asp Leu Thr Asp Ala Glu Trp
100 105 110

Phe Phe Leu Val Cys Met Ser Phe Ile Phe Asn Ile Gly Gln Gly Leu
115 120 125

Pro Gly Arg Thr Leu Ala Arg Asn Gln Ala Val Trp Leu Cys Asn Ala
130 135 140

His Arg Ala Asp Thr Lys Val Phe Ser Arg Ser Leu Leu Ala Lys Ser
145 150 155 160

Ala Ser Ile Gln Thr Val Val Cys Phe Pro Tyr Ser Glu Gly Val Val
165 170 175

Glu Leu Gly Ala Thr Glu Leu Val Pro Glu Asp Leu Asn Leu Ile Gln
180 185 190

His Ile Lys Thr Ser Phe Leu Asp Ser Pro Ala Thr Val Pro Lys Ile
195 200 205

Pro Asn Tyr Val Ser Asn Ser Ile Thr Asn Asn Asn Asp Leu Ile Cys
210 215 220

Glu Ala Leu Glu His Ala Asn Ile Pro Glu Asn Asp Leu Asp Gln Leu
225 230 235 240

```
Leu Asn Cys Pro Asp Thr Asn Ile Cys Ser Pro Asp Asn Ser Leu Asp
                245                 250                 255

Asp Phe Ala Asp Asn Leu Leu Ile Asp Glu Ser Asn Leu Ala Glu Gly
            260                 265                 270

Ile Asn Gly Glu Val Pro Gln Thr Gln Ser Trp Pro Phe Met Asp Asp
        275                 280                 285

Ala Ile Ser Asn Cys Leu Asn Ser Ser Met Asn Ser Ser Asp Cys Ile
    290                 295                 300

Ser Gln Thr His Glu Asn Leu Glu Ser Phe Ala Pro Leu Ser Asp Gly
305                 310                 315                 320

Lys Gly Pro Pro Glu Thr Asn Asn Cys Met His Ser Thr Gln Lys Cys
                325                 330                 335

Asn Gln Gln Ile Glu Asn Thr Gly Val Gln Gly Asp Glu Val His Tyr
            340                 345                 350

Gln Gly Val Leu Ser Asn Leu Leu Lys Ser Ser His Gln Leu Val Leu
        355                 360                 365

Gly Pro Tyr Phe Arg Asn Gly Asn Arg Glu Ser Ser Phe Val Ser Trp
    370                 375                 380

Asn Lys Asp Gly Ser Ser Gly Thr His Val Pro Arg Ser Gly Thr Ser
385                 390                 395                 400

Gln Arg Phe Leu Lys Lys Val Leu Phe Glu Val Ala Arg Met His Glu
                405                 410                 415

Asn Ser Arg Leu Asp Ala Gly Lys Gln Lys Gly Asn Ser Asp Cys Leu
            420                 425                 430

Ala Lys Pro Thr Ala Asp Glu Ile Asp Arg Asn His Val Leu Ser Glu
        435                 440                 445

Arg Lys Arg Arg Glu Lys Ile Asn Glu Arg Phe Met Ile Leu Ala Ser
    450                 455                 460

Leu Val Pro Ser Gly Gly Lys Val Asp Lys Val Ser Ile Leu Asp His
465                 470                 475                 480

Thr Ile Asp Tyr Leu Arg Gly Leu Glu Arg Lys Val Asp Glu Leu Glu
                485                 490                 495

Ser Asn Lys Met Val Lys Gly Arg Gly Arg Glu Ser Thr Thr Lys Thr
            500                 505                 510

Lys Leu His Asp Ala Ile Glu Arg Thr Ser Asp Asn Tyr Gly Ala Thr
        515                 520                 525

Arg Thr Ser Asn Val Lys Lys Pro Leu Thr Asn Lys Arg Lys Ala Ser
    530                 535                 540
```

```
Asp Thr Asp Lys Ile Gly Ala Val Asn Ser Arg Gly Arg Leu Lys Asp
545                 550                 555                 560

Ser Leu Thr Asp Asn Ile Thr Val Asn Ile Thr Asn Lys Asp Val Leu
                565                 570                 575

Ile Val Val Thr Cys Ser Ser Lys Glu Phe Val Leu Leu Glu Val Met
            580                 585                 590

Glu Ala Val Arg Arg Leu Ser Leu Asp Ser Glu Thr Val Gln Ser Ser
        595                 600                 605

Asn Arg Asp Gly Met Ile Ser Ile Thr Ile Lys Ala Lys Cys Lys Gly
    610                 615                 620

Leu Lys Val Ala Ser Ala Ser Val Ile Lys Gln Ala Leu Gln Lys Val
625                 630                 635                 640

Thr Met Lys Ser
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mouse myogenin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
Asp Arg Arg Arg Ala Ala Thr Leu Arg Glu Lys Arg Arg Leu Lys Lys
1               5                   10                  15

Val Asn Glu Ala Phe Glu Ala Leu Lys Arg Ser Thr Leu Asn Pro Asn
            20                  25                  30

Gln Arg Leu Pro Lys Val Glu Ile Leu Arg His Ala Ile Gln Tyr Ile
        35                  40                  45

Glu Arg Leu Gln Ala Leu
    50
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Yeast Cbf1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gln Arg Lys Asp Ser His Lys Glu Val Glu Arg Arg Arg Arg Glu Asn
1 5 10 15

Ile Asn Thr Ala Ile Asn Val Leu Ser Asp Leu Ile Pro Val Arg Glu
20 25 30

Ser Ser Lys Ala Ala Ile Leu Ala Arg Ala Ala Glu Tyr Ile Gln Lys
35 40 45

Leu Lys Glu Thr
50

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: AP-4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Ile Arg Arg Glu Ile Ala Asn Ser Asn Glu Arg Arg Arg Met Gln Ser
1               5                  10                  15

Ile Asn Ala Gly Phe Gln Ser Leu Lys Thr Leu Ile Pro His Thr Asp
            20                  25                  30

Gly Glu Lys Leu Ser Lys Ala Ala Ile Leu Gln Gln Thr Ala Glu Tyr
        35                  40                  45

Ile Phe Ser Leu Glu Gln Glu
    50                  55
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: N-myc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Glu Arg Arg Arg Asn His Asn Ile Leu Glu Arg Gln Arg Arg Asn Asp
1               5                  10                  15

Leu Arg Ser Ser Phe Leu Thr Leu Arg Asp His Val Pro Glu Leu Val
            20                  25                  30

Lys Asn Glu Lys Ala Ala Lys Val Val Ile Leu Lys Lys Ala Thr Glu
        35                  40                  45

Tyr Val His Ser Leu Gln Ala Glu
    50                  55
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: L-myc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Thr Lys Arg Lys Asn His Asn Phe Leu Glu Arg Lys Arg Arg Asn Asp
1 5 10 15

Leu Arg Ser Arg Phe Leu Ala Leu Arg Asp Gln Val Pro Thr Leu Ala
20 25 30

Ser Cys Ser Lys Ala Pro Lys Val Val Ile Leu Ser Lys Ala Leu Glu
35 40 45

Tyr Leu Gln Ala Leu Val Gly Ala
50 55

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Human E3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Gln Lys Lys Asp Asn His Asn Leu Ile Glu Arg Arg Arg Arg Phe Asn
1 5 10 15

Ile Asn Asp Arg Ile Lys Glu Leu Gly Thr Leu Ile Pro Lys Ser Ser
20 25 30

Asp Pro Glu Met Arg Trp Asn Lys Gly Thr Ile Leu Lys Ala Ser Val
35 40 45

Asp Tyr Ile Arg Lys Leu Gln Lys Glu
50 55

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Maize R-S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Ser Ala Thr Lys Asn His Val Met Ser Glu Arg Lys Arg Arg Glu Lys
1 5 10 15

Leu Asn Glu Met Phe Leu Val Leu Lys Ser Leu Leu Pro Ser Ile His
20 25 30

Arg Val Asn Lys Ala Ser Ile Leu Ala Glu Thr Ile Ala Tyr Leu Lys
35 40 45

Glu Leu Gln Arg Arg
50

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Maize Lc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Thr Gly Thr Lys Asn His Val Met Ser Glu Arg Lys Arg Arg Glu Lys
1 5 10 15

Leu Asn Glu Met Phe Leu Val Leu Lys Ser Leu Leu Pro Ser Ile His
20 25 30

Arg Val Asn Lys Ala Ser Ile Leu Ala Glu Thr Ile Ala Tyr Leu Lys
35 40 45

Glu Leu Gln Arg Arg
50

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Maize B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Asn Gly Ala Lys Asn His Val Met Ser Glu Arg Lys Arg Arg Glu Lys
1 5 10 15

Leu Asn Glu Met Phe Leu Val Leu Lys Ser Leu Val Pro Ser Ile His
20 25 30

Lys Val Asp Lys Ala Ser Ile Leu Ala Glu Thr Ile Ala Tyr Leu Lys
35 40 45

Glu Leu Gln Arg Arg
50

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Antirrhinum majus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Glu Ile Asp Arg Asn His Val Leu Ser Glu Arg Lys Arg Arg Glu Lys
1 5 10 15

Ile Asn Glu Arg Phe Met Ile Leu Ala Ser Leu Val Pro Ser Gly Gly
20 25 30

Lys Val Asp Lys Val Ser Ile Leu Asp His Thr Ile Asp Tyr Leu Arg
35 40 45

Gly Leu Glu Arg Lys
50

CLAIMS

1. A method of regulating the expression of one or more anthocyanin pigment genes in a plant which comprises the steps of transforming plant tissue with an expression vector comprising a DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 and which is capable of regulating expression of one or more plant genes involved in pigment biosynthesis, the said DNA segment being under the control of a promoter upstream of and operably linked thereto and regenerating from the transformed tissue plants showing altered anthocyanin pigmentation.

2. A method as claimed in claim 1 wherein the plant belongs to the dicotyledoneae.

3. A method as claimed in claim 1 or 2 wherein the DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 is a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2.

4. A plant having a DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 incorporated into its genome or plant propagation material of such a plant.

5. A plant as claimed in claim 4 wherein the plant belongs to the dicotyledoneae.

6. A plant as claimed in claim 4 or 5 wherein the DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 is a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2.

7. A DNA molecule encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO's 1 or 2 or a protein having an amino acid sequence which is at least 80%, preferably at least 90%, more preferably at least 98% similar with the *DEL* protein as shown in SEQ ID NO 1 or 2 and which is capable of regulating expression of one or more plant genes involved in pigment biosynthesis.

8. Use of the DNA molecule as claimed in claim 7 or the protein encoded thereby to isolate a DNA molecule encoding a protein having the amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 from other plant species.

9. A construct which comprises a transposon having cloned therein a DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2, the said DNA segment being under the control of a minimal promoter upstream of and operably linked thereto.

10. A method of trapping a promoter/enhancer which comprises the steps of transforming a plant with the construct as claimed in claim 9 and deriving from the transformed plant further plants having a phenotype showing altered anthocyanin pigmentation.

11. A method for isolating a trapped promoter/enhancer from a plant which has been transformed with the construct as

claimed in claim 9 which comprises reisolating the construct from said plant together with sequences adjacent thereto.

12. A method of expressing a gene of interest in a plant, which comprises transforming a cell of said plant with a first construct having said gene of interest under the control of a first promoter, which first promoter is that of an anthocyanin gene, upstream of and operably linked thereto, the said plant having incorporated into its genome a DNA segment encoding a protein having the amino acid sequence of the *DEL* protein as shown in SEQ ID NO 1 or 2 or a protein having an amino acid sequence which shows substantial homology with the *DEL* protein as shown in SEQ ID NO 1 or 2 under the control of a second promoter upstream of and operably linked thereto, or the said plant being co-transformed with a second construct which comprises said DNA segment under the control of a third promoter, which third promoter may be the same or different to the second promoter, upstream of and operably linked thereto, or the said first construct optionally including the said DNA segment under the control of said second or third promoter upstream of and operably linked thereto if the said plant does not have incorporated into its genome the said DNA segment or is not co-transformed with the said second construct, and deriving from the transformed plant further plants expressing said gene of interest.

13. The method as claimed in claim 12 wherein the said second promoter for the said DNA segment is a promoter/enhancer isolated by the method as claimed in claim 11.

14. A method of expressing a gene of interest in a plant which comprises transforming said plant with the construct as claimed in claim 8, deriving from the transformed plant further plants having a phenotype showing altered anthocyanin pigmentation, reisolating from said plant the said construct

together with sequences adjacent thereto, replacing said DNA segment in said construct with a gene of interest to form a new construct and transforming said plant with said new construct.

FIG. 1(1)

FIG. 1(1)	FIG. 1(2)
FIG. 1(3)	FIG. 1(4)

```
   1  gtagagaggagagaggattcaagaatggctactggtacccaaaaccaaaaagata
   1                                M  A  T  G  I  Q  N  Q  K  I

 124  tggtccaattcagttgcacaaccaggggtcttggagtggggtgatgggttctac
  34  W  S  N  S  V  A  Q  P  G  V  L  E  W  G  D  G  F  Y

 247  agaagtgatcaattgagagaactttatgagtctctttcacttggtgaaaccaac
  75  R  S  D  Q  L  R  E  L  Y  E  S  L  S  L  G  E  T  N

 370  gtttgcatgtctttcatattcaatattggccaagggttgcctggaagaacatta
 116  V  C  M  S  F  I  F  N  I  G  Q  G  L  P  G  R  T  L

 493  cttgcaaagagtgcgtcaattcagacagttgtgtgctttccatattcagaaggt
 157  L  A  K  S  A  S  I  Q  T  V  V  C  F  P  Y  S  E  G

 616  ttcttggacagtcctgccaccgttcccaagattcccaactatgtctccaacagt
 198  F  L  D  S  P  A  T  V  P  K  I  P  N  Y  V  S  N  S

 739  cagcttttgaattgtccagacacgaacatatgttctcctgataacagtttggat
 239  Q  L  L  N  C  P  D  T  N  I  C  S  P  D  N  S  L  D

 862  acacaaagctggcctttcatggatgatgcaatcagcaattgtctcaatagttct
 280  T  Q  S  W  P  F  M  D  D  A  I  S  N  C  L  N  S  S
```

FIG. 1(2)

```
gtgcctgagaatttgaggaaycaacttgctattgctgtgagaagtatccaatggagttatgcaattttc
V  P  E  N  L  R  K  Q  L  A  I  A  V  R  S  I  Q  W  S  Y  A  I  F

aatggagatattaaaactcgaaaaactgtacaatctgtcgaattgaatcaagatcagctgggattgcag
N  G  D  I  K  T  R  K  T  V  Q  S  V  E  L  N  Q  D  Q  L  G  L  Q

acacaagctaaaaggcctactgctgcattatcaccagaagacctcactgatgctgagtggtttttcttg
T  Q  A  K  R  P  T  A  A  L  S  P  E  D  L  T  D  A  E  W  F  F  L

gcacgaaatcaagcagtatggctatgcaacgctcatcgtgcggacaccaaagttttctcgcgttctttg
A  R  N  Q  A  V  W  L  C  N  A  H  R  A  D  T  K  V  F  S  R  S  L

gtagttgagctgggagcaacagagctagtaccggaggatttgaatctaatccagcatataaaaacttca
V  V  E  L  G  A  T  E  L  V  P  E  D  L  N  L  I  Q  H  I  K  T  S

attacaaacaacaatgacctcatttgtgaagcgcttgaacatgctaatataccagaaaacgatcttgat
I  T  N  N  N  D  L  I  C  E  A  L  E  H  A  N  I  P  E  N  D  L  D

gactttgcagacaatttactcatagacgaatcgaatttggcagaaggcatcaatggggaggttcctcaa
D  F  A  D  N  L  L  I  D  E  S  N  L  A  E  G  I  N  G  E  V  P  Q

atgaattctagtgactgtatatctcaaactcatgaaaatctagagtcttttgctccactttctgatgga
M  N  S  S  D  C  I  S  Q  T  H  E  N  L  E  S  F  A  P  L  S  D  G
```

FIG. 1(3)

```
 985  aaagggccaccggagacgaataattgtatgcacagcactcaaaaatgcaatcag
 321  K  G  P  P  E  T  N  N  C  M  H  S  T  Q  K  C  N  Q

1108  agttcccatcagttggttcttggtccctacttcagaaatgggaatagagaatca
 362  S  S  H  Q  L  V  L  G  P  Y  F  R  N  G  N  R  E  S

1231  tttctgaagaaagtacttttgaagtagctagaatgcatgaaaactccaggctt
 403  F  L  K  K  V  L  F  E  V  A  R  M  H  E  N  S  R  L

1354  cacgtcttgtcagagagaaaacgcagagagaaaataaacgaacggtttatgatt
 444  H  V  L  S  E  R  K  R  R  E  K  I  N  E  R  F  M  I

1477  ttgagagggcttgagaggaaagtcgacgagctggaatctaacaaaatggtaaag
 485  L  R  G  L  E  R  K  V  D  E  L  E  S  N  K  M  V  K

1600  ggcgcaacaaggacaagtaacgtcaagaaaccgttgacaaacaagagaaaggct
 526  G  A  T  R  T  S  N  V  K  K  P  L  T  N  K  R  K  A

1723  actgtgaacattacaaacaaggatgtgttgattgtcgtgacttgttcttccaag
 567  T  V  N  I  T  N  K  D  V  L  I  V  V  T  C  S  S  K

1846  tccaacagagatggaatgatatctattaccataaaagccaagtgcaagggattg
 608  S  N  R  D  G  M  I  S  I  T  I  K  A  K  C  K  G  L

1969  atgctcactatctatagctagcttttgtgtaaaaaatttgtattcataacttttt
```

FIG. 1(4)

```
cagatagaaaacacgggtgtccaaggcgatgaggtccattatcaaggggtactttccaatcttttgaag
Q  I  E  N  T  G  V  Q  G  D  E  V  H  Y  Q  G  V  L  S  N  L  L  K

agcttcgttagttggaacaaggatggatcgtcgggtactcatgttccccgaagcggaacctcacaaaga
S  F  V  S  W  N  K  D  G  S  S  G  T  H  V  P  R  S  G  T  S  Q  R

gatgctggtaaacaaaagggcaacagtgactgccttgcaaagccaacggctgatgaaattgatagaaac
D  A  G  K  Q  K  G  N  S  D  C  L  A  K  P  T  A  D  E  I  D  R  N

cttgcatccctagtcccatccggtggcaaggttgacaaagtatcaatactagaccatacaatagattac
L  A  S  L  V  P  S  G  G  K  V  D  K  V  S  I  L  D  H  T  I  D  Y

ggccgggggcgggaatcaactacaaaaactaaactacacgatgccattgagaggacctctgataattat
G  R  G  R  E  S  T  T  K  T  K  L  H  D  A  I  E  R  T  S  D  N  Y

tctgatacggacaagattggagccgtaaatagcagaggtcgattgaaagattccttaacagataatata
S  D  T  D  K  I  G  A  V  N  S  R  G  R  L  K  D  S  L  T  D  N  I

gagtttgtattgcttgaagtgatggaagccgtaagacgactaagtttggattccgaaactgttcaatct
E  F  V  L  L  E  V  M  E  A  V  R  R  L  S  L  D  S  E  T  V  Q  S

aaggttgcatcagcaagtgtgatcaaacaagctcttcagaaagttactatgaagtcttgaagttgattt
K  V  A  S  A  S  V  I  K  Q  A  L  Q  K  V  T  M  K  S  *

gctaagtaatttgcagggcttttccaagtagttcagatcaataaaaaaaaaaa
```

FIG. 2A

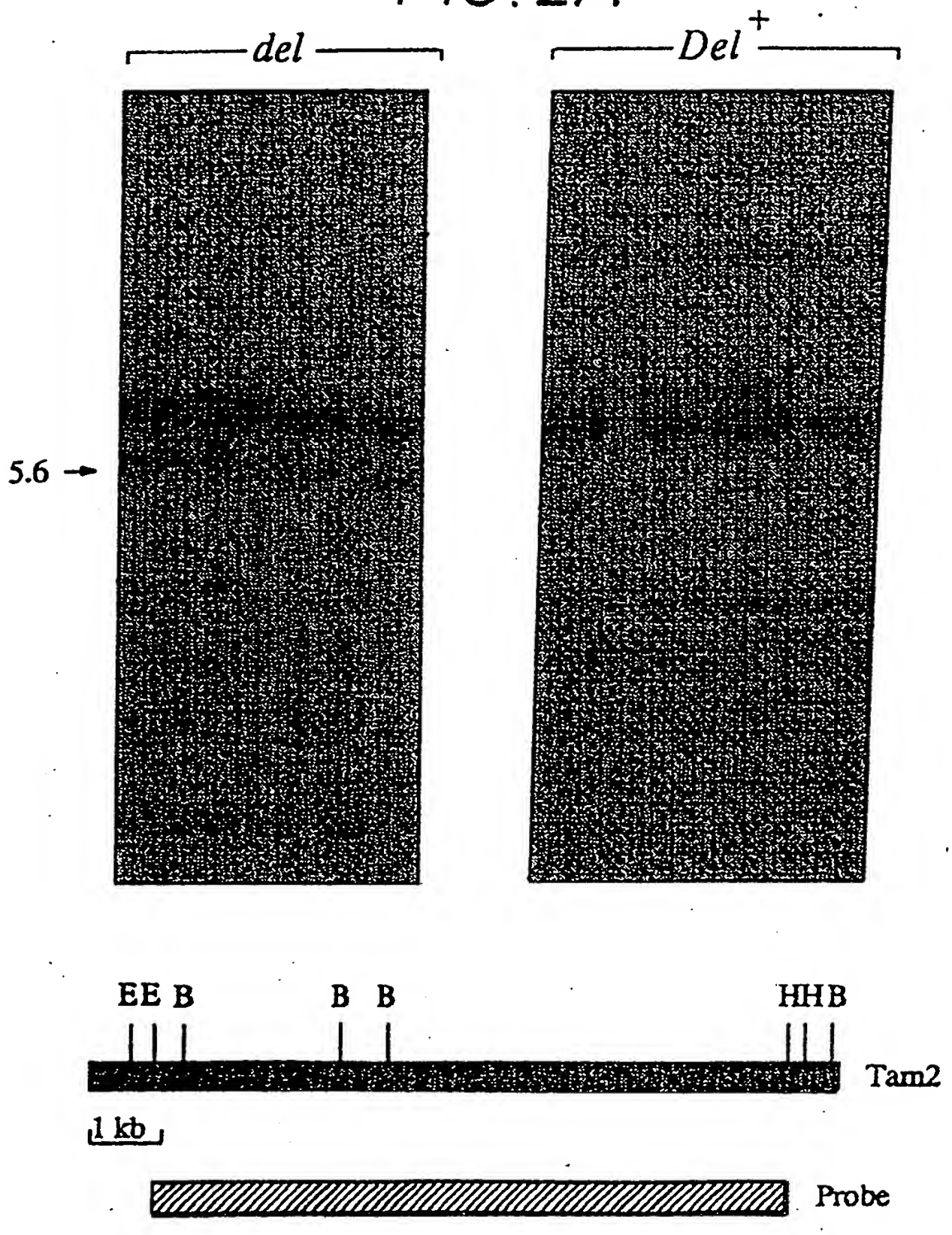


FIG. 2B

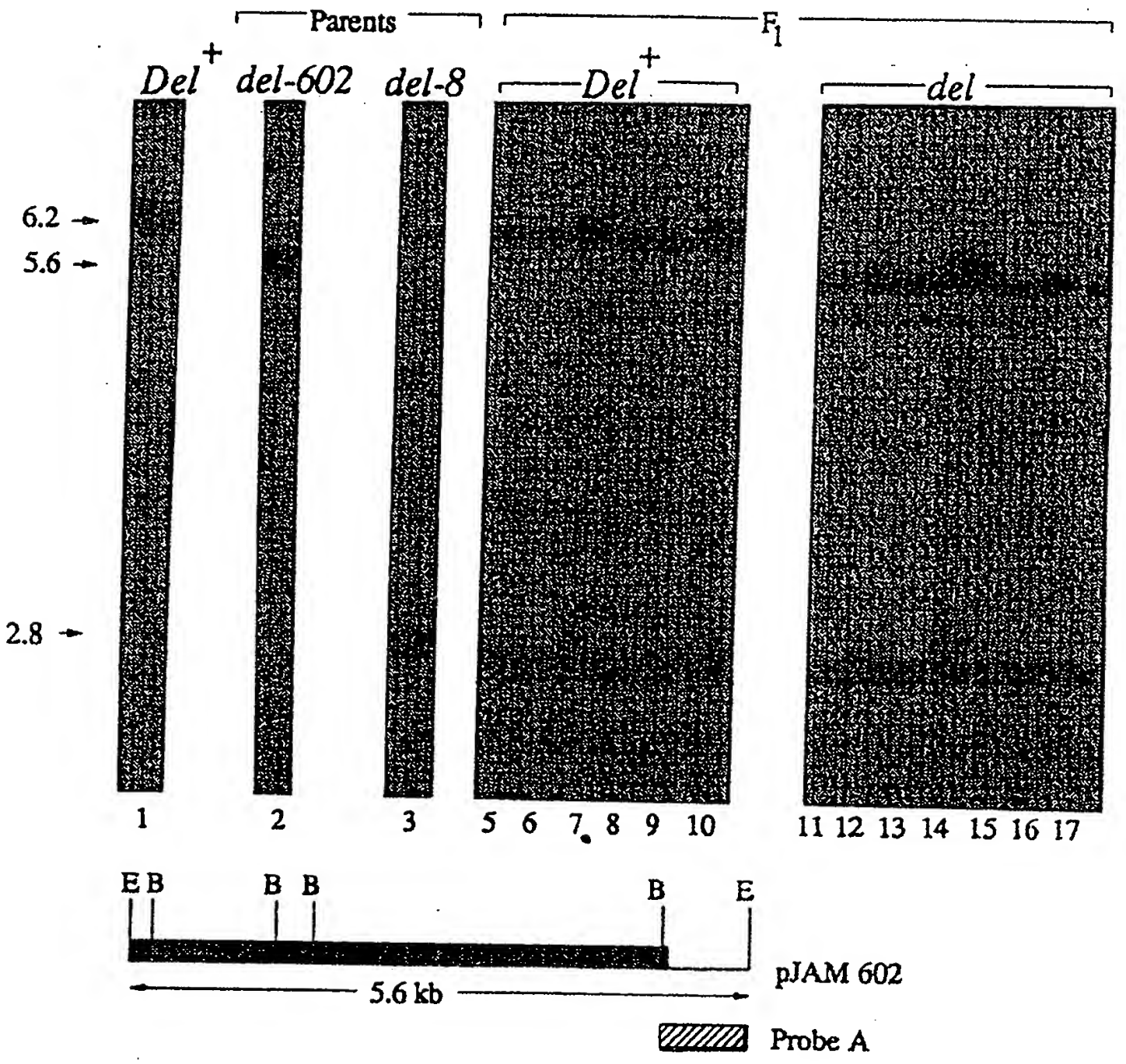
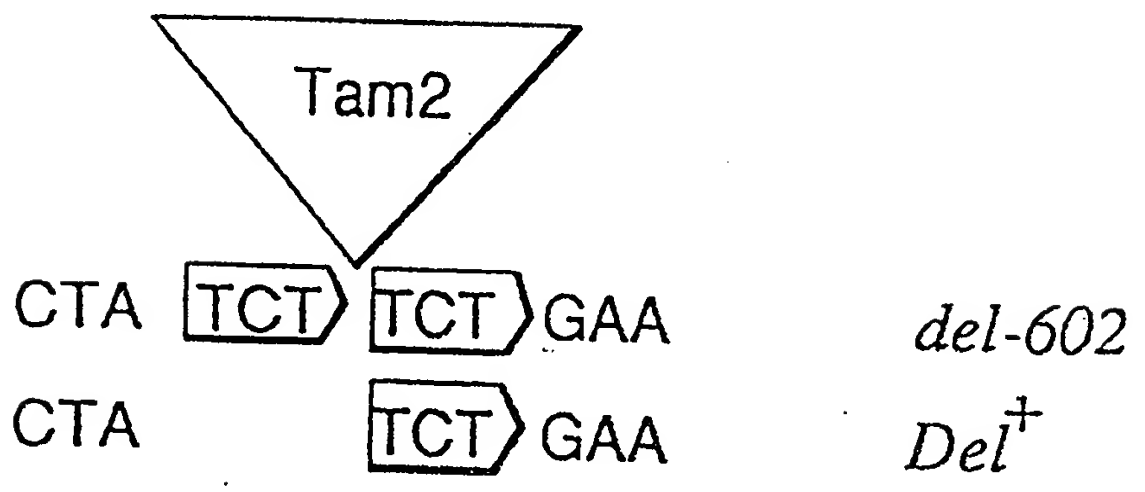

Parents
F1
Del+
del-602
del-8
Del+
del
6.2
5.6
2.8
1
2
3
5 6 7 8 9 10
11 12 13 14 15 16 17
E B
B B
B
E
pJAM 602
5.6 kb
Probe A

FIG. 3


Tam2
CTA TCT TCT GAA
del-602
CTA TCT GAA
Del+

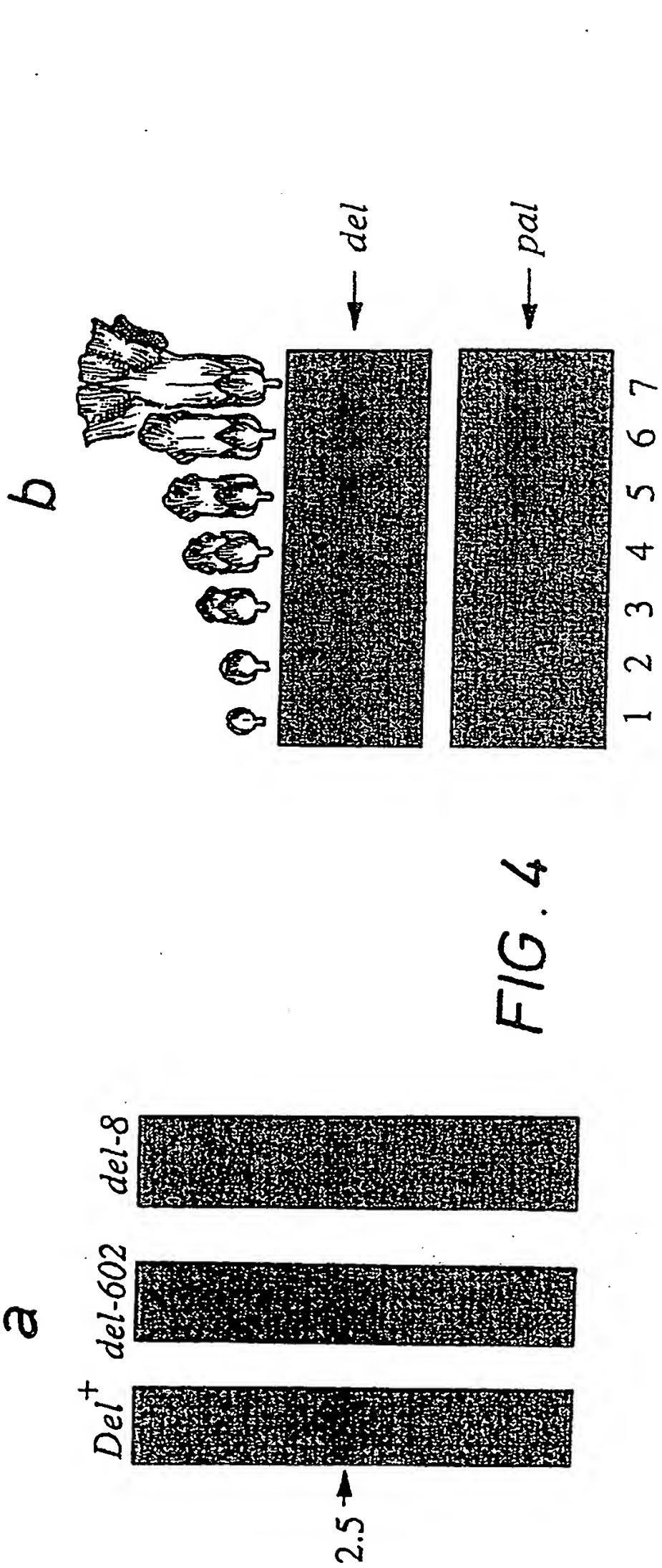

a
Del+
del-602
del-8
2.5
b
del
pal
1 2 3 4 5 6 7
c
BT RT L
del
d
L S B Se P St Sty C
del

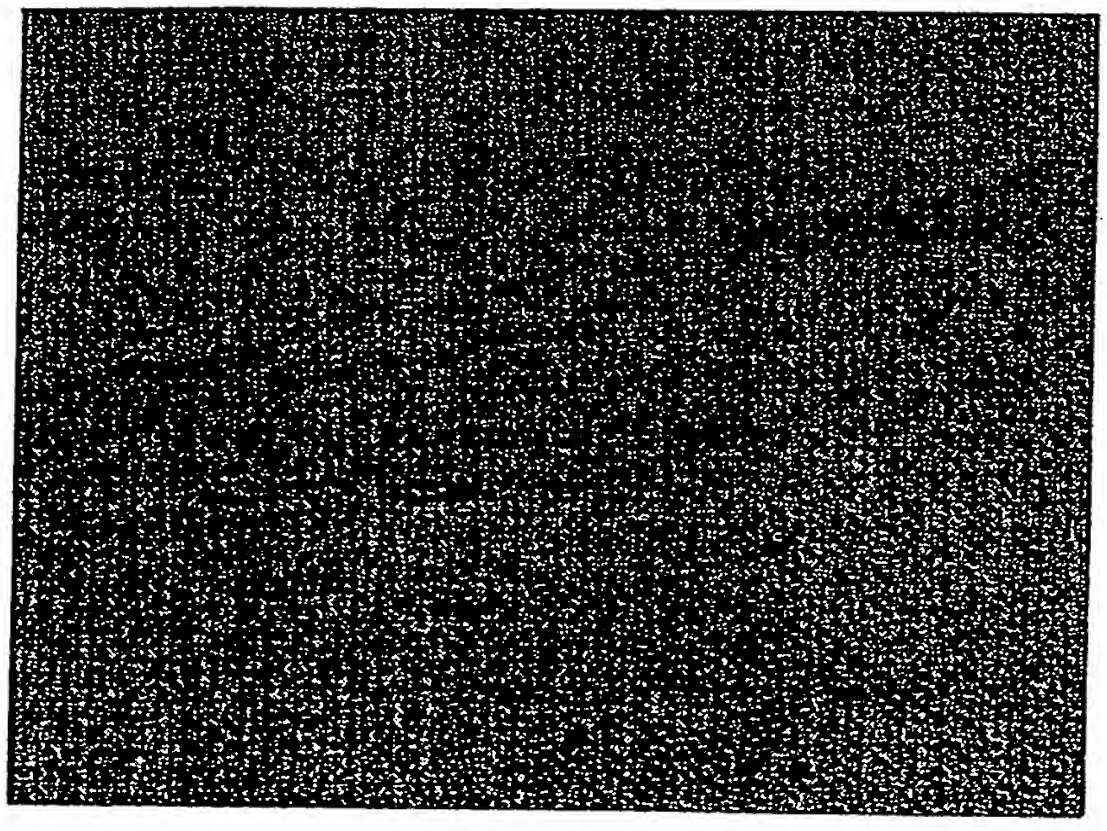
FIG. 4

FIG. 5

	BASIC DOMAIN / HELIX I / LOOP	HELIX II
Myogenin	DRRRAATLREKRRLKKVNEAFEALKRSTLNPN	QRLPKVEILRHAIQYIERLQAL
Cbf1	QRKDSHKEVERRRRENINTAINVLSDLIPVR	ESSKAAILARAAEYIQKLKET
AP-4	IRREIANSNERRRMQSINAGFQSLKTLIPHTDGE	KLSKAAILQQTAEYIFSLEQE
N Myc	ERRRNHNILERQRRNDLRSSFLTLRDHVPELVKNE	KAAKVVILKKATEYVHSLQAE
L Myc	TKRKNHNFLERKRRNDLRSRFLALRDQVPTLASC$	KAPKVVILSKALEYLQALVGA
E3	QKKDNHNLIERRRRFNINDRIKELGTLIPKSSDPE	MRWNKGTILKASVDYIRKLQKE
R-S	SATKNHVMSERKRREKLNEMFLVLKSLLPSIH	RVNKASILAETIAYLKELQRR
Lc	TGTKNHVMSERKRREKLNEMFLVLKSLLPSIH	RVNKASILAETIAYLKELQRR
B	NGAKNHVMSERKRREKLNEMFLVLKSLVPSIH	KVDKASILAETIAYLKELQRR
DEL	EIDRNHVLSERKRREKINERFMILASLVPSGG	KVDKVSILDHTIDYLRGLERK
HLH family consensus	RR N ER R ψN F LK ψP	K K IL Aψ Yψ LQ
	KK R	R VE

8/11

FIG. 6a



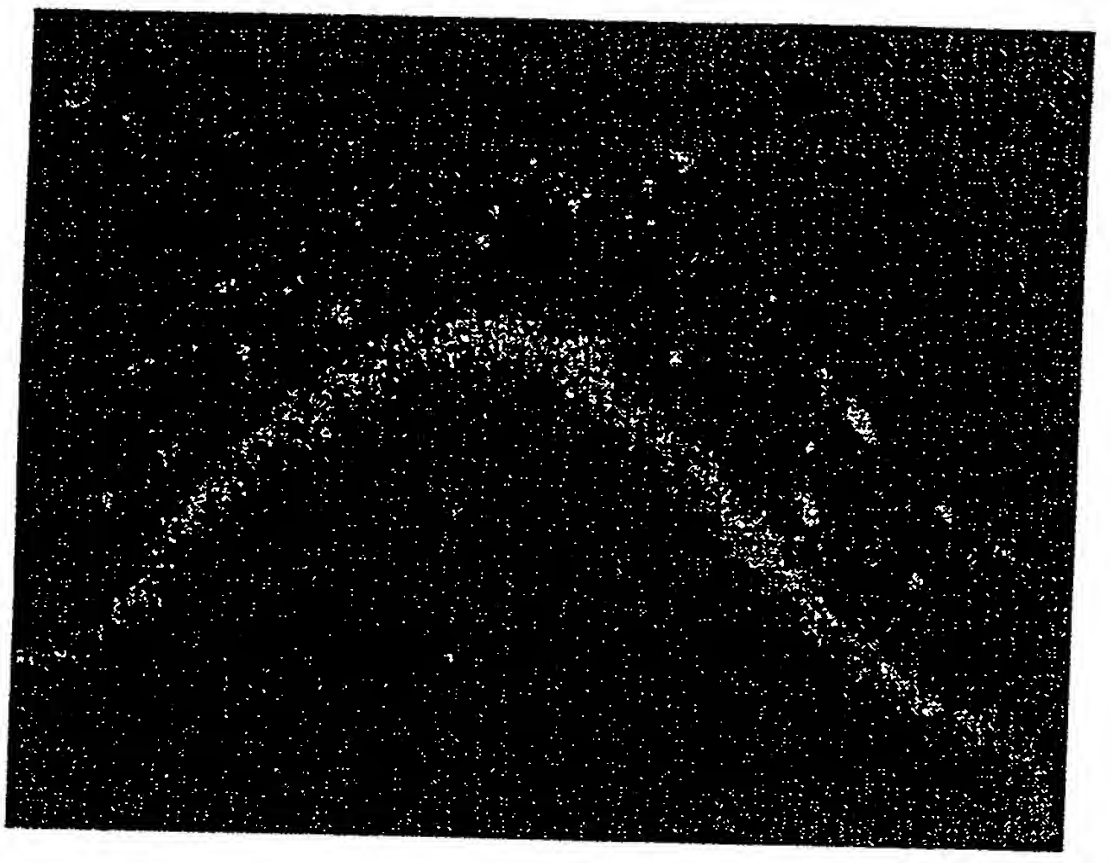


FIG. 6b

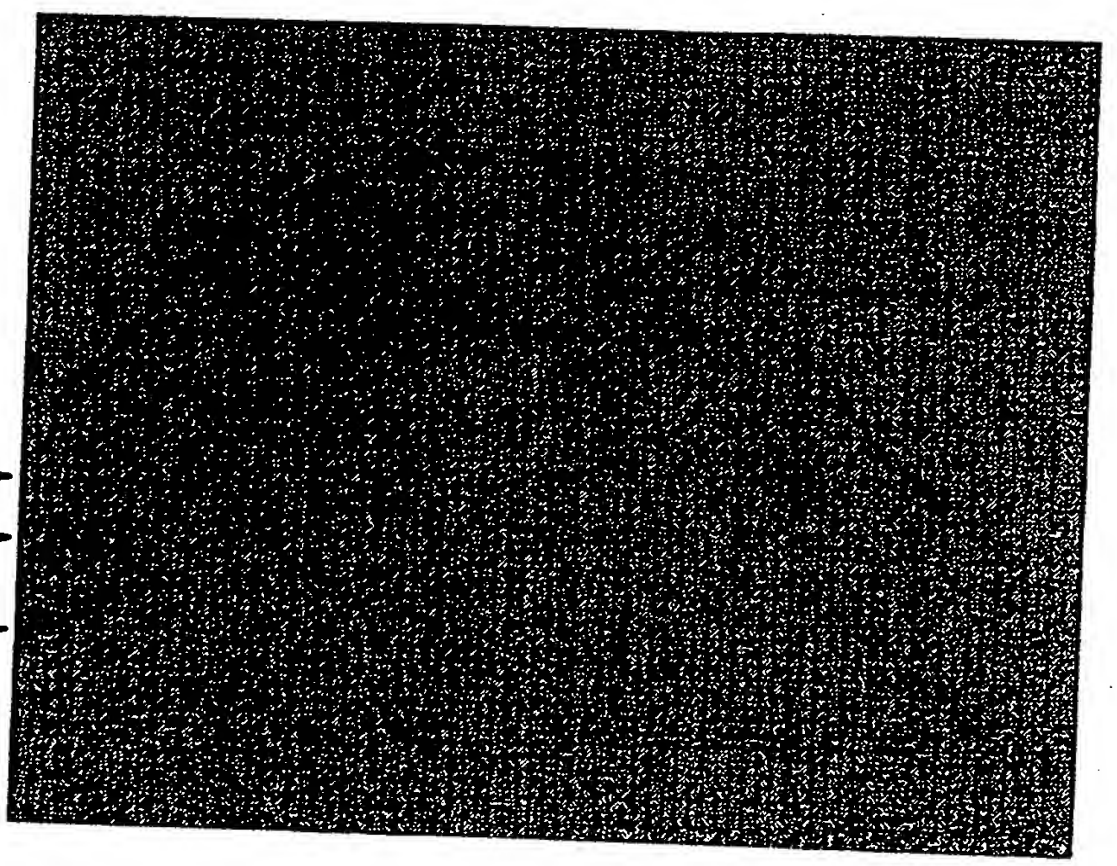
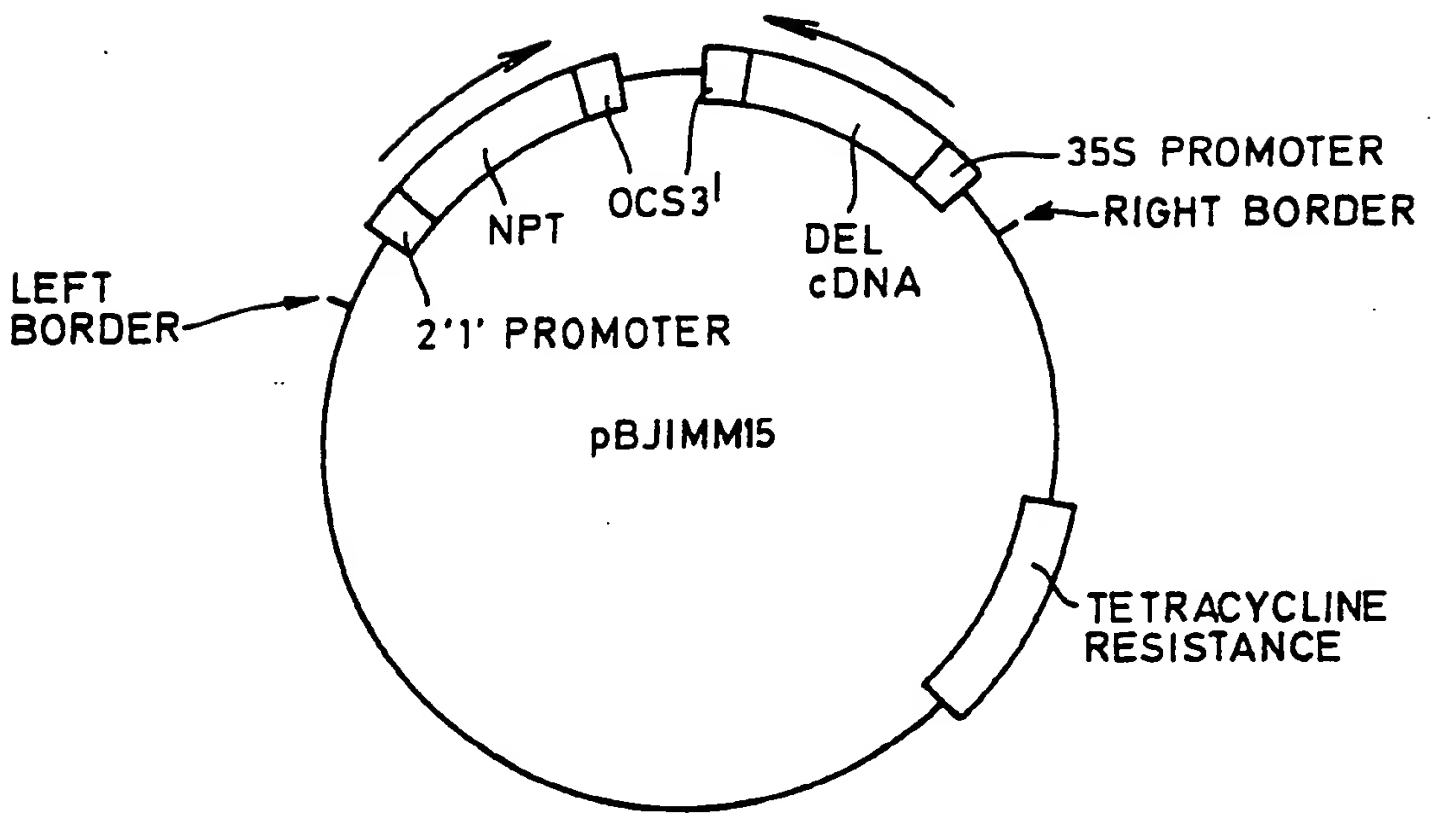
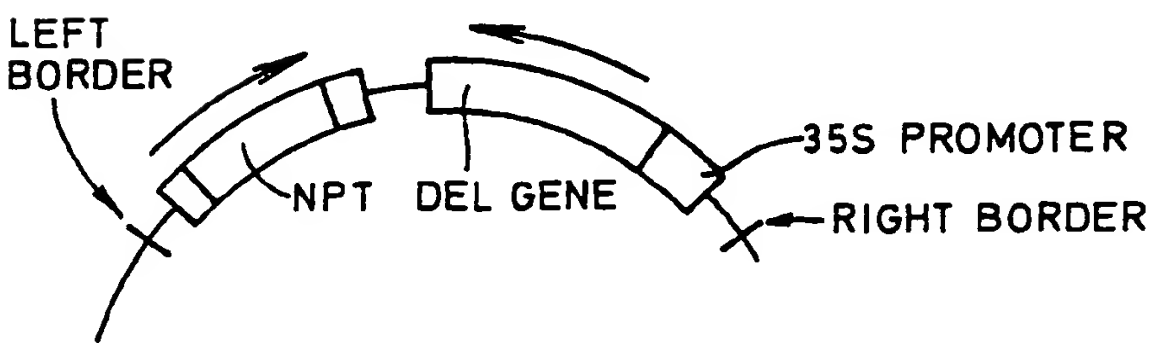
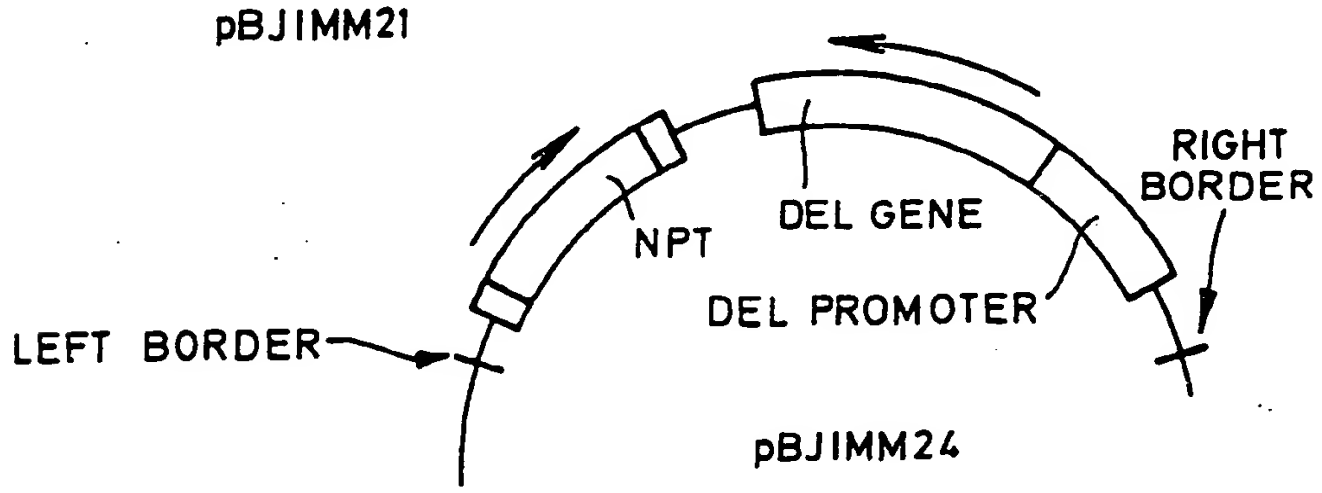


FIG. 7





pBJIMM21



pBJIMM24

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 93/00019

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/82; C12N15/29; C12Q1/68; A01H5/00
A01H5/02

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C12Q ; A01H

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
X	J. CELL. BIOCHEM. SUPPL. vol. 15A, 1991, page 114 GOODRICH, J.W., ET AL. 'Molecular cloning and characterisation of Delila, a gene regulating the anthocyanin biosynthesis pathway in Antirrhinum majus' see abstract A514	7
Y	---	1-6,12
Y	WO,A,9 102 059 (PIONEER HI-BRED) 21 February 1991 see claim 20	1-6

	-/--	

° Special categories of cited documents :[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
12 MAY 1993	03. 06. 93
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	MADDOX A.D.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	J. CELL. BIOCHEM. SUPPL. vol. 15A, 1991, page 132 JACKSON, D.P., ET AL. 'Expression patterns of structural and regulatory genes in Antirrhinum flowers' see abstract A614	1-6
Y	EMBO JOURNAL vol. 9, no. 8, 1990, EYNSHAM, OXFORD GB pages 2517 - 2522 GOFF, S.A., ET AL. 'Transactivation of anthocyanin biosynthetic genes following transfer of B regulatory genes into maize tissues' see the whole document	12
O,P, X	PLANT MOLECULAR BIOLOGY. vol. 19, 1992, DORDRECHT, THE NETHERLANDS. pages III - VI LIFSCHITZ, E. 'News and views' see page IV, paragraph 3 & WORKSHOP, THE MOLECULAR CONTROL OF FLOWER DEVELOPMENT AND PLANT REPRODUCTION, HELD APRIL 1992, AMSTERDAM.	1-7
P,X	CELL vol. 68, 6 March 1992, CAMBRIDGE, NA US pages 955 - 964 GOODRICH, J., ET AL. 'A common gene regulates pigmentation pattern in diverse plant species' see the whole document	7
A	THE PLANT JOURNAL vol. 1, no. 1, 1991, pages 37 - 49 MARTIN, C., ET AL. 'Control of anthocyanin biosynthesis in flowers of Antirrhinum majus' cited in the application see the whole document	1-14

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

GB 9300019
SA 68549

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 12/05/93

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9102059	21-02-91	AU-A- 6287390	11-03-91
		EP-A- 0462231	27-12-91